



Modélisation et simulation de l'interdépendance entre l'objet, l'observateur et le modèle de l'objet dans la Triade de Minsky. Application à la surveillance épidémiologique en santé animale.

Bruno Bonté

► To cite this version:

Bruno Bonté. Modélisation et simulation de l'interdépendance entre l'objet, l'observateur et le modèle de l'objet dans la Triade de Minsky. Application à la surveillance épidémiologique en santé animale.. Modélisation et simulation. Université Montpellier II - Sciences et Techniques du Languedoc, 2011. Français. NNT: . tel-00719426

HAL Id: tel-00719426

<https://theses.hal.science/tel-00719426>

Submitted on 19 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

Public Domain

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ MONTPELLIER II

N° attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--

THÈSE

pour obtenir le grade de

Docteur de l'Université Montpellier 2

Spécialité : **Informatique**

préparée aux laboratoires **Animal et Gestion Intégrée des Risques (AGIRs)**
et Gestion des Ressources Renouvelables (GREEN) du CIRAD

dans le cadre de l'École Doctorale **ISS**

présentée et soutenue publiquement
par

Bruno Bonté

le 16 décembre 2011

Titre:

**Modélisation et simulation de l'interdépendance entre
l'objet, l'observateur et le modèle de l'objet dans la Triade
de Minsky.
Application à la surveillance épidémiologique en santé
animale.**

Directeur de thèse: **Jean-Pierre Müller**

Co-encadrant de thèse: **Raphaël Duboz**

Jury

Pr. Benoit Durand,	Président du jury
Dr. Mamadou Kaba Traoré,	Rapporteur
Pr. Frédérick Garcia,	Rapporteur
Dr. Fabien Michel,	Examineur
Dr. François Roger,	Invité
Pr. Jean-Pierre Müller,	Directeur de thèse
Dr. Raphaël Duboz,	Co-encadrant de thèse

*Il ne concevait pas de savoir véritable
auquel ne correspondit pas quelque pouvoir
d'action. Créer, construire étaient pour lui
inséparable de connaître et comprendre.*
P. Valéry, Préface aux Cahiers de Léonard de Vinci, 1942

À mes parents,

Remerciements

Je remercie Jean-Pierre Müller d'avoir accepté de diriger cette thèse et de m'avoir toujours encouragé dans mon travail.

Je ne saurais comment décrire la participation de Raphaël Duboz à ce travail tant ma proposition de thèse me semble émerger d'elle même de l'ensemble des discussions que nous avons eu et des lectures qu'il m'a conseillées. Quoiqu'il en soit, je le remercie vivement pour son accompagnement, son investissement et son amitié. Il a finalement su me pousser à poser une véritable question de recherche en informatique et à y apporter une réponse personnalisée.

Je remercie Frédérick Garcia, Directeur de Recherches et responsable du département Mathématiques et Informatique Appliquées à l'INRA, d'avoir accepté d'être rapporteur de cette thèse. Son rapport positif est un honneur pour moi et les conseils qu'il a pu me donner au cours de ma thèse furent précieux. Je remercie également Mamadou Kaba Traoré, Maître de Conférences HDR à l'Université Blaise Pascal à Clermont Ferrand, d'avoir accepté d'être rapporteur de ma thèse. Je n'ai pas eu la chance de pouvoir travailler avec lui mais une partie de ma thèse est inspirée de ses travaux. Je considère que son rapport, à la fois enthousiaste et critique, est une chance pour moi et pour ce travail de thèse.

Je remercie grandement Benoit Durand, Directeur de Recherche à l'ANSES d'avoir accepté de faire partie de mon jury de thèse, il réunit la vision de l'épidémiologiste, du vétérinaire et du modélisateur qu'il était essentiel d'avoir dans mon jury.

Je remercie enfin Fabien Michel, Chargé de Recherches au LIRMM, pour sa participation à ce jury. Cette thèse a été inspirée par une rencontre entre des idées de l'Intelligence Artificielle et la modélisation des systèmes complexes et je suis content de pouvoir avoir son regard sur ce travail. L'optimisme dont il a fait part lors de ma soutenance de thèse m'encourage grandement.

Je remercie doublement François Roger. Pour avoir accepté mon invitation à la soutenance de cette thèse et pour m'avoir laissé la liberté nécessaire à la réalisation d'une thèse en informatique au sein de l'équipe AGIRs. Son ouverture m'a permis de me sentir légitime dans la position parfois délicate de modélisateur informaticien dans une équipe focalisée sur la gestion des risques liée à la santé animale.

J'en arrive ainsi à remercier l'ensemble de l'équipe AGIRs pour son accueil chaleureux et pour toutes les interactions que j'ai eu avec les uns et les autres et qui m'ont permis d'apercevoir les enjeux de la surveillance épidémiologique en santé animale. Je me dois par honnêteté professionnelle de remercier Timothée Vergne et Mathilde Paul qui ont participé à la construction des modèles à travers les projets que l'on avait en commun ; et Julien Cappelle et Stéphanie Desveaux pour s'être penchés avec attention et intérêt sur la définition de la triade "cas d'école" qui fonde

la partie applicative de ma thèse. Travailler avec eux fut un plaisir productif. J'aimerais de plus les inclure dans les remerciements personnels que j'aimerais également adresser à mes secrétaires blonde et brune, respectivement Carlène et Ève, les Marie et Marie-Anne, les membres des ateliers débat du midi avec Vladimir, Lucas, Véro, Nicolas, ...

Je transmets également de chaleureux remerciements à l'ensemble de l'équipe GREEN qui m'a accueilli au CIRAD bien avant ma thèse, forgeant ma vision de la modélisation et de la recherche-action au service des sociétés. Un merci tout particulier à François, Christophe, Martine et Aurélie pour l'attention qu'ils ont portés tout au long de cette thèse à mon travail ... et à mes états d'âmes dans les moments de stress. Merci pour votre soutien chaleureux et inconditionnel. Un clin d'œil également à mes compagnons de thèse Fany, Paolo, Panomsak et Manuela.

Je remercie les membres de mon comité de thèse, Frédéric Garcia et François Bousquet déjà cités, ainsi que Agnes Varet, Éric Ramat et Alexis Drogoul. J'ai eu la chance qu'ils se penchent sérieusement sur mon travail. Ils ont toujours suivi l'objectif de me guider vers une question qui m'intéresse et que je puisse traiter en fonction du temps qu'il me restait et des compétences auxquelles je pourrais avoir accès.

Merci également aux personnes qui m'ont accueillies au département d'informatique de l'AIT, les grands et gros Oliviers, O, et le groupe des thésards.

Un grand merci également aux membres du salon # vle (mon ami comme dirait Eric Ramat). Je n'aurais jamais pu programmer les simulateurs dont j'avais besoin sans l'aide de Gauthier Quesnel, Bl0b, Eric', Pierre Casa et les autres membre de l'équipe de développement de VLE.

Enfin un merci général à mes proches, amis et famille, qui ont également participé (plus ou moins directement..) à cette aventure.

Résumé

On ne peut pas reproduire l'expérience d'une épidémie à l'échelle d'un pays ou d'une région. Or, en l'absence d'expérience reproductible, la notion usuelle de « validation » de modèle qui consiste à tester si le modèle A^* d'un système A permet de reproduire le comportement de A , n'a aucun sens statistique. Marvin Minsky donne la définition suivante de ce qu'est un modèle : Pour un observateur B , un objet A^* est un modèle d'un objet A s'il permet à B de répondre à une question qu'il se pose sur A . Nous appelons triade de Minsky, l'ensemble des trois objets A , B et A^* . Nous proposons d'utiliser la Théorie de la Modélisation et de la Simulation (TMS) pour modéliser et simuler la triade de Minsky. Cela nous permet de modéliser la triade vue comme un système dynamique composé des objets A , B et A^* . Nous pouvons ainsi nous interroger sur l'utilisation d'un modèle A^* par un utilisateur B et sur l'impact que cette utilisation a sur la trajectoire du système A . Nous appliquons ce cadre à une triade de Minsky empruntée à notre contexte d'étude. Il s'agit d'un cas d'école où une épidémie (l'objet A) est observée et contrôlée par un système de surveillance et de contrôle (l'observateur B) et où un modèle épidémiologique (A^*) est utilisé pour évaluer les mesures de contrôle. Grâce à cet exemple nous pouvons discuter comment le cadre de modélisation et de simulation de la triade de Minsky permet aux chercheurs de se poser de nouvelles questions et aux décideurs de mieux comprendre l'impact des modèles qu'ils utilisent.

Abstract

The spread of a disease at national or international scale isn't a reproducible experiment. If experiment is not reproducible, the usual concept of model validation has no statistical meaning because it would involve the comparison between model behavior and system behavior. We use Marvin Minsky's definition of model : « to an observer B, an object A^* is a model of an object A to the extent that B can use A^* to answer questions that interest him about A ». The three objects A, B and A^* are the triad of Minsky. We propose to use the Theory of Modelling and Simulation (TMS) to model and simulate the triad seen as a dynamic system composed of the objects A, B and A^* . We hence can answer questions about the use of A^* and the impact it has on A. We apply this framework to a triad of Minsky in epidemiological surveillance in animal health. An epidemics (object A) is observed and controlled by a surveillance system and a control system (observer B) and an epidemiological model (model A^*) is used to evaluate the control measures. Thanks to this application, we can discuss how our framework enables researchers to formulate new questions in Modelling and Simulation and decision makers to better understand their use of Models.

Table des matières

Remerciements	v
Résumé	vii
Abstract	viii
Table des matières	ix
Table des figures	xiii
Acronymes	xv
Notations	xvii
 Avant-Propos	 1
 Introduction	 3
 1 Problématique	 7
1.1 Introduction	7
1.2 Contexte thématique	8
1.2.1 L'épidémiologie	8
1.2.2 La surveillance épidémiologique en santé animale	9
1.3 Les modèles et leur utilisation	12
1.3.1 Les modèles	12
1.3.2 L'utilisation des modèles	14
1.4 Problématique	18
1.4.1 Positionnement de la thèse en informatique	21
1.5 Plan de la thèse	22
 2 État de l'art	 23
2.1 Introduction	23
2.1.1 Théorie systémique et cybernétique	24
2.1.2 La cognition et le paradoxe de « l'observation - action »	25
2.2 A : Les systèmes épidémiologiques, sociologiques et écologiques	26
2.2.1 Les systèmes complexes	26
2.2.2 Résilience des systèmes sociologiques et écologiques	28
2.2.3 Les Socio-Éco-Systèmes	28
2.3 A_B^* : Les modèles	29
2.3.1 Modèles de représentation de l'incertain	29
2.3.2 Les modèles de systèmes dynamiques	29
2.3.3 Le paradigme multi-agent	30
2.3.4 Les modèles de modèles	31
2.4 ρ_o : l'observation et le contrôle de systèmes complexes	33
2.4.1 Les modèles statistiques	33

2.4.2	Les outils statistiques d'estimation de prévalence	34
2.4.3	Approches statistiques pour le monitoring et la surveillance . .	35
2.5	ρ_m : L'utilisation des modèles	36
2.5.1	Résultats analytiques	36
2.5.2	Comprendre pour contrôler	37
2.5.3	Les cadres expérimentaux	37
2.5.4	Calibration	38
2.5.5	Évaluation des systèmes de surveillance par simulation	39
2.6	B : Processus de décision	39
2.6.1	Théorie de la décision	39
2.7	Cadre général : la théorie de la modélisation et de la simulation . . .	40
2.7.1	Systèmes et niveaux de spécification	40
2.7.2	Les entités de base de la TMS et leurs relations	41
2.7.3	La validation et la vérification	44
2.7.4	Une sémantique opérationnelle : DEVS	45
2.7.5	Conclusion : Retour sur la problématique	49
3	Proposition	51
3.1	Introduction : le modèle du système Triade	51
3.1.1	T : Le système triade	52
3.1.2	T* : Le modèle du système triade	53
3.1.3	Plan du chapitre	55
3.2	A_C^* : Le modèle de système cible	57
3.3	ρ_m^* : Le modèle de processus expérimental	58
3.3.1	Un modèle d'expérimentation de modèles	58
3.3.2	Définition d'un protocole expérimental itératif	59
3.3.3	Modélisation du processus expérimental avec DEVS	61
3.3.4	Exemples d'utilisation d'un MPE	64
3.4	A^{**} : Le modèle A^{**} cadré	65
3.5	ρ_o^* : Le modèle de SPOC	66
3.5.1	Les modèles du système d'observation	66
3.5.2	Les modèles du système de contrôle	72
3.5.3	Le modèle de système cible contrôlé	72
3.6	B* : Le modèle de processus de décision	73
3.7	Implémentation et expérimentation de T* avec VLE	74
4	Application	77
4.1	Introduction : La triade T_{uc}'	77
4.2	Système « <i>under control</i> » (T_{uc})	78
4.2.1	Situation d'épidémie contrôlée	78
4.2.2	Système T_{uc}	79
4.3	Questions sur le système T_{uc}	80
4.4	Modèle « <i>under control</i> » (T_{uc}^*)	82
4.4.1	Le modèle de système cible	83
4.4.2	Le modèle de SPOC	85
4.4.3	Le modèle de processus expérimental	88
4.4.4	Modèle de processus de décision	91
4.5	Expérimentation de T_{uc}^*	91

4.5.1	Une expérience du modèle T_{uc}^*	92
4.5.2	Plan d'expérience	94
4.5.3	Analyse des résultats d'expérience	97
4.6	Enseignements sur le système T_{uc}	100
4.6.1	Réponses aux questions sur le système T_{uc}	100
4.6.2	Nouvelles questions sur T_{uc}	101
4.7	Conclusion	102
5	Discussion	103
5.1	Introduction : une contribution en trois outils	103
5.2	Des outils à enrichir	105
5.2.1	Le cadre conceptuel T'	105
5.2.2	Le modèle formel T^*	106
5.2.3	L'implémentation modulaire de T^* dans VLE	108
5.3	Des outils à utiliser	108
5.3.1	Application à la surveillance épidémiologique en santé animale	108
5.3.2	Application aux systèmes sociologiques et écologiques	109
5.3.3	Application à l'étude des systèmes dynamiques en général	111
6	Conclusion	117
6.1	Un cadre méthodologique pour l'évaluation des modèles	117
6.1.1	Le cadre conceptuel T'	118
6.1.2	La spécification formelle du modèle T^*	119
6.2	Limites et perspectives d'utilisation	119
6.2.1	Un cadre d'évaluation réflexif et relatif	119
6.2.2	Des pistes concrètes pour la surveillance épidémiologique et la gestion des Socio-Éco-Système (SES)	120
6.2.3	Un point de vue constructiviste	121
	Bibliographie	123
	Annexes	133
A	The Virtual Laboratory Environment	135
A.1	Classe de modèle	135
A.2	Graph translator	135
B	Niveaux de spécifications de systèmes	137
B.1	Cadre d'observation d'entrée/sortie (niveau 0)	137
B.2	Relation d'entrée/sortie (niveau 1)	140
B.3	Fonction d'entrée/sortie (niveau 2)	141
B.4	Système d'entrée/sortie (niveau 3)	142
B.5	Spécification itérative de système (niveau 4)	143
B.6	Système structuré (niveau 5)	144
B.7	Système multicomposants (niveau 6)	144
B.8	Systèmes couplés structurés (niveau 7)	145

C	Modèles composants de T*	147
C.1	Modèle d'observation passive retardée	147
C.2	Modèle d'observation planifiée	147
D	MPE d'estimation de paramètres utilisant une méthode d'optimisation par essaim de particules	149
E	Approximation par moments d'un modèle épidémiologique individu-centré	155
F	Le Modèle de Processus Expérimental : Application à la modélisation multi-échelles	183
G	Vers une spécification des modèles de simulation de systèmes complexes	199
H	Résultats de simulation	233
I	Exemples d'utilisation du modèle de système épidémiologique	235
I.1	Simulation de la propagation de la grippe aviaire hautement pathogène dans le réseau de production du Nord Vietnam	235
I.2	Simulation de méthodes de captures-recaptures comme un nouvel outil d'évaluation de la surveillance en santé animale	239
J	Utilisation d'un modèle de système épidémiologique couplé à un modèle de surveillance épidémiologique pour l'aide à la conception des systèmes de surveillance	243

Table des figures

1	Organisation du document	5
1.1	Domaines de l'épidémiologie abordés dans la thèse	9
1.2	Les quatre niveaux institutionnels d'un réseau de surveillance épidé- miologique (d'après [DH07]).	10
1.3	La triade de Minsky	14
1.4	La recherche en informatique appliqué à la M&S dans la triade de Minsky	15
1.5	Démarche inductive	15
1.6	Domaines de l'épidémiologie et modèles	17
1.7	Cadre général : La triade T'	21
1.8	Organisation de la thèse autour de la triade T'	22
2.1	La Cognition et le paradoxe de l'observation - action de R. Vallée . .	26
2.2	Illustration d'un attracteur d'un système SIS	38
2.3	Notion de Cadre Expérimental	42
2.4	Hiérarchie de systèmes couplés	48
2.5	Simulateur hiérarchique	49
3.1	Modélisation et simulation de la triade T*	52
3.2	Le système Triade	54
3.3	Structure d'un Modèle T*	56
3.4	Principe de simulation encapsulée (d'après [GS05])	59
3.5	Protocole expérimental itératif	61
3.6	Structure du modèle de processus expérimental. (Image généré par le logiciel VLE.)	62
3.7	Diagramme d'états DEVS du modèle exécutif du système cible contrôlé.	64
3.8	Modèle A** cadré	65
3.9	Schéma de connexion entre un modèle d'observation passive et un modèle composant du système cible.	68
3.10	Diagramme de phase DEVS d'un modèle d'observation passive	68
3.11	Schéma de connexion entre un modèle d'observation proactive et un modèle composant du système cible.	69
3.12	Diagramme de phase DEVS d'un modèle d'observation proactive . . .	70
3.13	Schéma de connexion entre un modèle d'observation réactive et un modèle composant du système cible.	71
3.14	Diagramme de phase DEVS d'un modèle d'observation réactive . . .	71
3.15	Modèle exécutif du modèle de système cible contrôlé.	76

3.16	Diagramme de phase DEVS d'un modèle exécutif de système cible contrôlé	76
4.1	La triade T_{uc}'	78
4.2	Chronologie des mesures de surveillance et de prophylaxie dans la situation d'épidémie contrôlée.	80
4.3	Structure du modèle T_{uc}^*	84
4.4	Modèle de propagation de la maladie	85
4.5	Simulation de propagation de la maladie	86
4.6	Schéma des connexions entre les modèles composant le modèle de Système Pour l'Observation et le Contrôle (SPOC) de T_{uc}^*	88
4.7	Cadre expérimental et modèle utilisés par le MPE de T_{uc}^*	90
4.8	Sorties de simulations du modèle T_{uc}^*	93
4.9	Illustration du calcul des critères c_1 à c_4 à partir des sorties de simulation de T_{uc}^* (cf. section 4.5.1).	95
4.10	Relations entre les critères	101
5.1	Une contribution à trois étage	104
5.2	Montée en abstraction de Minsky	112
5.3	Cadre infini de la connaissance de Minsky	113
B.1	Dynamique de multiplication et d'excrétion d'un agent pathogène infectieux par un hôte (superposition de graphiques de [TDS ⁺ 08]) . . .	137
B.2	Système « Hôte d'une maladie contagieuse » IO	138
B.3	Le système « hôte » : IOR	141
B.4	Le système « hôte » : IOF	142
B.5	Hierarchie de systèmes couplés	146
C.1	Schéma de connexion entre un modèle d'observation delayée et un modèle composant du système cible.	147
C.2	Schéma de connexion entre un modèle d'observation planifiée et un modèle composant du système cible.	148
D.1	Déplacement d'une particule dans un algorithme OEP	151
D.2	Cadre expérimental pour l'estimation de paramètres	152

Acronymes

CE Cadre Expérimental

DESS Differential Equations System Specification

DEVS Discrete Event System Specification

DS-DEVS Dynamic Structure Discrete Event System Specification

DTSS Discrete Time System Specification

MADC Monitoring, Analyse, Décision, Contrôle

MPE Modèle de Processus Expérimental

MSD Modèle de Système Dynamique

M&S Modélisation et Simulation

ODD Overview, Design concepts, Detail

OEP Optimisation par Essaims de Particules

OIE Office Internationale des Épizooties

PE Processus Expérimental

QSS Quantized System Specification

REVASIA Recherches pour l'EValuation en Asie du sud est de la Surveillance de
l'Influenza Aviaire

SCS Society for Computer Simulation

SES Socio-Éco-Système

SIRS Susceptible, Infectieux, Rétablit, Susceptible

SPOC Système Pour l'Observation et le Contrôle

TMS Théorie de la Modélisation et de la Simulation

VLE Virtual Laboratory Environment

Notations

T Triade de Minsky

A Objet dans une triade de Minsky

B Observateur dans une triade de Minsky

A* Modèle dans une triade de Minsky

ρ_o Relation entre l'observateur et l'objet dans la triade Minsky

ρ_m Relation entre l'observateur et le modèle dans la triade Minsky

T* Modèle de triade de Minsky

T' Triade de Minsky dont l'objet est lui-même une triade (T) et le modèle est un modèle de triade (T*)

T_{uc} Exemple de triade en épidémiologie animale : la triade « under control »

Avant-Propos

Un environnement pluri-disciplinaire

Cette thèse a été réalisée au CIRAD (centre de Coopération Internationale en Recherche Agronomique pour le Développement) et plus précisément en co-encadrement entre les équipes AGIRs (Animal et Gestion Intégrée des Risques) et GREEN (Gestion des Ressources Renouvelables et Environnement). Le CIRAD est un organisme de recherche français qui s'intéresse principalement aux problématiques des pays du sud. En conséquence, les travaux de recherches des équipes AGIRs et GREEN concernent des problématiques que l'on peut trouver dans des pays en voie de développement.

De manière générale, les recherches menées au CIRAD, et particulièrement les recherches menées dans le département « Environnement et Sociétés » du CIRAD, sont des recherches qui sont tournées vers l'action et qui font intervenir de nombreuses disciplines. Un des enjeux du département « Environnement et Sociétés » (dont font partie les équipes GREEN et AGIRs) est de mettre en place des cadres d'études interdisciplinaires qui permettent une analyse rigoureuse du développement des sociétés humaines et de leur interaction avec leur environnement naturel (écosystèmes, faune, flore, ..) et anthropisé (territoires, agriculture, élevage, économie, ...).

Indications pour la lecture du manuscrit

Ce manuscrit de thèse a été rédigé de manière à développer une argumentation qui aille de l'introduction à la discussion. La courte introduction présente comment cette argumentation est développée en spirale chapitre après chapitre. Une lecture linéaire du manuscrit est conseillée mais le chapitre 3, très technique, peut être survolé par un lecteur non-informaticien. À l'inverse, un lecteur ayant une bonne connaissance de la Théorie de la Modélisation et de la Simulation (TMS) peut se contenter de lire le chapitre 3 qui propose un modèle formel générique. Nous l'invitons cependant à lire le chapitre de discussion qui développe les enseignements que nous avons tirés.

Le chapitre de conclusion a été écrit pour être auto-suffisant et reprendre l'ensemble des résultats obtenus.

Introduction

Ce travail de thèse s'inscrit en informatique et plus particulièrement dans le domaine de la Modélisation et Simulation (M&S). L'objectif était de formuler des questions génériques et d'y apporter des réponses génériques. Néanmoins, les réflexions qui nourrissent la thèse ont eu lieu dans le cadre des travaux effectués sur les problématiques de l'équipe AGIRs ou de l'équipe GREEN (durant les trois années qu'a duré la thèse et les quelques années qui précèdent). Ces trois années de thèse en informatique se sont écoulées dans un souci constant de généralisation des questions posées par les thématiques de AGIRs ou de GREEN et auxquelles la M&S pouvait éventuellement apporter des réponses.

D'un point de vue thématique, la thèse se positionne sur les questions de recherche de l'équipe AGIRs et plus particulièrement sur la surveillance épidémiologique des maladies infectieuses en santé animale.

Cette problématique propose de nombreuses questions de recherche dans lesquelles la M&S intervient. Le sujet initial de cette thèse s'intitulait par exemple « Modélisation et simulation multi-agents appliquées à la gestion des risques épidémiques ». Nous sommes montés en abstraction en nous interrogeant sur l'utilisation de la Modélisation et de la Simulation dans les contextes étudiés. Nos réflexions découlent de la problématique suivante : « Quelle validité des modèles en l'absence d'expérience reproductible » ? Cette question faisait déjà partie intégrante des questions de recherche de l'équipe GREEN. Nous espérons avoir trouvé une manière originale et pertinente de l'aborder pour le cas des modèles de systèmes dynamiques¹. Nous proposons de considérer le triplet « objet, modèle et observateur » (l'observateur observe l'objet et construit et utilise le modèle). Nous appelons ce triplet la « triade de Minsky » en référence à la définition de modèle donnée par Marvin Minsky et qui fait intervenir ces trois entités. Dans cette thèse, nous construisons un modèle de cette triade pour étudier la relation entre les entités de cette triade. La thématique de la surveillance épidémiologique des maladies infectieuses en santé animale a constitué le contexte que nous avons utilisé pour discuter autour de cette problématique.

Une fois cette thèse identifiée, un des enjeux de ce travail de thèse a été d'expliquer clairement comment cette question de recherche en informatique, abstraite et générique, peut être instanciée sur une question rencontrée par un chercheur d'une autre discipline, et de montrer ainsi comment le cadre méthodologique et technique associé à la thèse peut être utilisé. Dans cette optique, le document est structuré sous la forme d'allers-retours entre les questions thématiques qui constituent le contexte qui amène la question de thèse, la question générique posée, et le cadre méthodolo-

1. Différentes acceptions du mot validité pour ce type de modèles sont données dans le chapitre d'état de l'art (section 2.7.3 page 44)

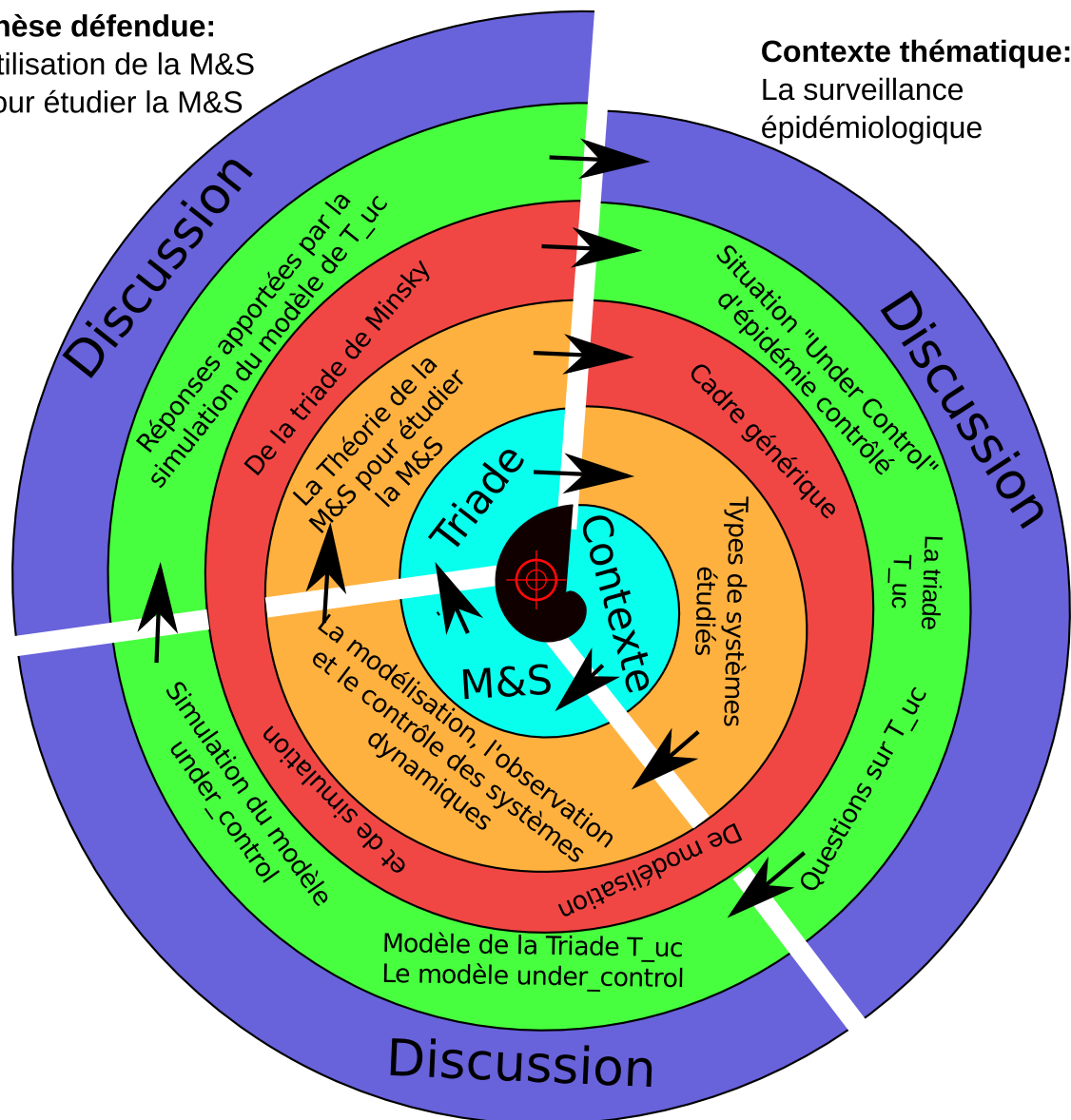
gique et technique proposé. On peut représenter l'argumentation de cette thèse sous la forme d'une spirale (figure 1). Un chapitre d'énoncé de la problématique présente brièvement : le contexte thématique (la surveillance épidémiologique des maladies infectieuses en santé animale) ; les domaines d'applications qui sont la modélisation et la simulation ; la place de ces recherches au sein de la M&S. Il introduit et justifie la question de recherche de la thèse. Un deuxième chapitre présente ensuite l'état de l'art. Il reformule la question de recherche de manière plus précise. Notre proposition de cadre méthodologique et technique est faite dans un troisième chapitre. Dans le quatrième chapitre une situation empruntée au domaine d'application est présentée et le cadre méthodologique et technique est utilisé dans cette situation. Le dernier chapitre discute enfin chacun des points abordés.

Thèse défendue:

Utilisation de la M&S
pour étudier la M&S

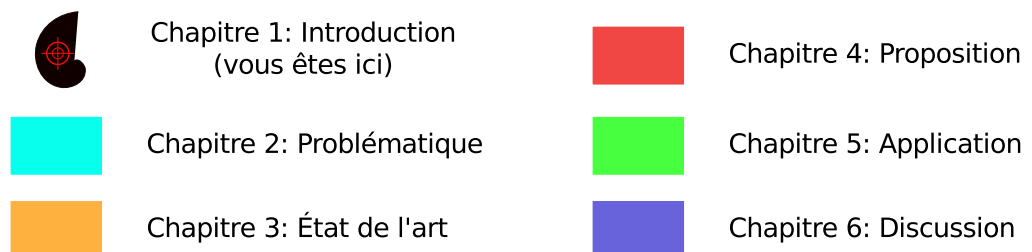
Contexte thématique:

La surveillance
épidémiologique



Domaines de recherche:

La modélisation et la simulation (M&S)



M&S = Modélisation et simulation

FIGURE 1: Organisation du document

Chapitre 1

Problématique

RÉSUMÉ

Notre contexte thématique est la surveillance épidémiologique des maladies infectieuses en santé animale et nos recherches en informatique portent sur la Modélisation et la Simulation M&S. Nous nous intéressons plus particulièrement dans cette thèse à l'utilisation de la M&S par les autres domaines de recherche, nous présentons donc cette utilisation en l'illustrant dans notre contexte thématique. Nous situons ensuite la place de l'informatique dans le domaine de la M&S. Ces préambules nous permettent d'expliquer comment le contexte thématique de la surveillance épidémiologique nous a amené à notre proposition de thèse en M&S. Cette thèse postule qu'il est possible et souhaitable d'utiliser la M&S pour le domaine de recherche de la M&S lui-même. Nous situons enfin notre travail de thèse par rapport à la M&S. Nous finissons en exposant le plan de la thèse.

1.1 Introduction

L'informatique se définit dans le langage courant comme le domaine d'activité scientifique, technique et industriel concernant le traitement automatique de l'information par des machines telles que : calculateur, système embarqué, ordinateur, console de jeux vidéo, robot, automate, etc.¹. Cette thèse s'inscrivant dans une spécialité informatique, elle s'intéresse davantage à l'informatique sous sa forme théorique qui serait définie comme une branche des mathématiques. Ceci dit, cette thèse se positionne en « informatique appliquée » par opposition à l'informatique théorique. De la même manière que les mathématiques appliquées vont s'intéresser à l'application de résultats mathématiques à d'autres domaines, nos questions de recherche vont porter sur les applications des sciences formelles qui composent l'informatique (logique, algorithmique, ...), davantage que sur ces sciences elles-mêmes. Notre domaine d'application est la M&S dans laquelle l'informatique joue un rôle

1. Première phrase de la page française de Wikipédia sur l'informatique le 16 octobre 2011 [Wik11]. Le dictionnaire hachette donne la définition suivante. « Informatique : Technique du traitement automatique de l'information au moyen des calculateurs et des ordinateurs. » [Fou98]

prépondérant. La M&S s'intéresse à la spécification de modèles de systèmes dynamiques et à la simulation de ces modèles par des ordinateurs. Il s'agit bien d'informatique appliqué. Ainsi, les outils informatiques que nous utilisons le plus dans cette thèse sont liés à la spécification formelle de processus dynamiques qui est définie dans la théorie de la modélisation et de la simulation par Bernard P. Zeigler [ZKP00]. Nous présentons la TMS et l'ensemble des méthodes informatiques mobilisées par la thèse dans le chapitre sur l'état de l'art.

La modélisation et la simulation s'appliquent à des contextes thématiques. On modélise forcément quelque chose, pour une certaine raison. De ces contextes émergent les défis adressés à la M&S, et auxquels nos propositions doivent répondre. Ces défis, que nous développons dans cette problématique, concernent l'utilisation de la M&S dans des situations où les expériences ne sont pas reproductibles et où l'observateur (celui qui construit et utilise le modèle) fait partie du système observé.

Dans une première section nous présentons le contexte thématique qui est la surveillance épidémiologique des maladies infectieuses en santé animale. Dans une seconde section nous présentons la modélisation et la simulation en insistant sur l'utilisation des modèles pour répondre à des questions thématiques et en positionnant l'informatique dans ces domaines de recherche. Dans une troisième section nous exposons la problématique que le contexte thématique nous semble adresser à la M&S. Il s'agit de la non-reproductibilité de l'expérience qui diminue la confiance que l'on peut avoir en les résultats apportés par les modèles. Nous exposons ensuite notre proposition qui doit permettre d'évaluer un modèle en modélisant l'utilisation qui en est faite. Enfin, dans une dernière partie nous exposons le plan de ce manuscrit qui doit nous permettre de défendre la thèse avancée dans notre problématique.

1.2 Contexte thématique

1.2.1 L'épidémiologie

L'épidémiologie est définie par les sciences médicales (humaines et vétérinaires) comme l'étude des maladies et des facteurs de santé dans une population [TDS⁺08]. Étymologiquement, le mot épidémiologie signifie « science de ce qui se rapporte à la population ». Il s'agit d'un domaine très vaste qui peut être découpé en plusieurs sous domaines. Nous distinguons principalement l'épidémiologie descriptive, l'épidémiologie analytique et l'épidémiologie théorique (cf. figure 1.1).

L'épidémiologie descriptive a pour but d'étudier la répartition des phénomènes de santé dans les populations. Elle mobilise un ensemble de modèles et d'outils statistiques et y associe un vocabulaire précis permettant de décrire les situations épidémiologiques. L'épidémiologie analytique utilise l'épidémiologie descriptive pour tenter d'établir des relations causales entre les phénomènes de santé et les comportements des pathogènes, des individus, ou leur environnement. Enfin, l'épidémiologie théorique étudie de manière abstraite dans une population, des processus d'émergence, de diffusion, de maintien ou d'extinction de caractéristiques individuelles éventuellement transmissibles d'un individu à un autre. C'est dans le domaine de l'épidémiologie théorique que sont définis les modèles épidémiologiques mathématiques, dont l'utilisation date des années 80 [AM79]. L'épidémiologie analytique est en fait une succession d'allers-retours entre l'épidémiologie descriptive et l'épidémio-

logie théorique.

La surveillance épidémiologique joue un rôle important puisqu'elle consiste à récolter les données de départ sur lesquelles reposent toutes les descriptions et donc les analyses.

Les domaines de l'épidémiologie	Rôle
Épidémiologie théorique	Étudier formellement et théoriquement des phénomènes d'émergence, de diffusion de maintien ou d'extinction de caractéristiques individuelles dans une population
Épidémiologie analytique	Expliquer de manière causale les phénomènes de santé observés à l'échelle de la population
Épidémiologie descriptive	Décrire de manière synthétique les phénomènes de santé à l'échelle de la population

FIGURE 1.1: Domaines de l'épidémiologie abordés dans la thèse

1.2.2 La surveillance épidémiologique en santé animale

Réseaux de surveillance et réseaux de contrôle

La surveillance épidémiologique formalise les processus d'observation des épidémies en caractérisant les activités liées à la collecte d'informations épidémiologiques sous la forme de dispositifs de surveillance épidémiologique. Les sciences médicales distinguent la prophylaxie qui consiste à lutter contre une maladie et la surveillance qui consiste à observer son comportement épidémiologique [DH07]. Pour faire un parallèle avec la médecine classique à l'échelle de l'individu, on pourrait dire que la surveillance épidémiologique regroupe les méthodes de diagnostic (c'est-à-dire la caractérisation d'un état clinique pouvant aider à l'action) à l'échelle d'une population et que la prophylaxie regroupe les méthodes de traitement de la population.

La surveillance épidémiologique est organisée sous forme de réseaux de surveillance. La structure des réseaux de surveillance épidémiologique en santé animale est schématisée figure 1.2. Cette organisation découpe le réseau en différents niveaux d'intervention [DH07] :

- 1^{er}- Niveau terrain : éleveurs regroupés en Groupements de Défense Sanitaire
- 2^{ème}- Étage local : vétérinaires praticiens qui peuvent effectuer des prélèvements (Groupement Technique Vétérinaire)
- 3^{ème}- Niveau intermédiaire : unités vétérinaires d'État et laboratoires provinciaux
- 4^{ème}- Niveau central

Cette organisation n'est pas strictement hiérarchique, ni dans le sens où les niveaux supérieurs auraient autorité sur les niveaux inférieurs, ni dans le sens où les niveaux inférieurs seraient systématiquement inclus dans les niveaux supérieurs. Mais

de manière générale, les niveaux inférieurs couvrent des régions géographiques, ou des parties de la population cible plus petites que les niveaux supérieurs. Par exemple le comité de pilotage décide des actions des postes de surveillance avec l'aide (l'éclairage) des unités intermédiaires et du comité technique alors que l'unité centrale gère la centralisation et les relations entre les intervenants, les équipes mobiles aident à la vie du réseau, et le niveau 1 décide individuellement de sa participation [DH07].

Le fonctionnement du réseau est basé sur la définition précise de cas de maladie que le réseau doit détecter. Les données (cas détectés) sont transmises des niveaux inférieurs aux niveaux supérieurs et éventuellement de manière transversale aux unités du même niveau d'intervention. En contre partie les informations agrégées et synthétisées sont redistribuées des niveaux supérieurs aux niveaux inférieurs.

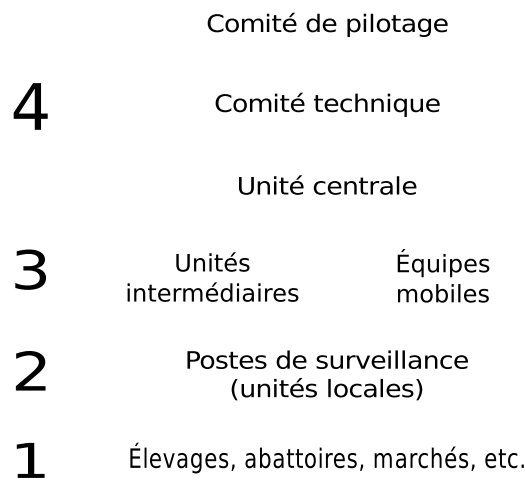


FIGURE 1.2: Les quatre niveaux institutionnels d'un réseau de surveillance épidémiologique (d'après [DH07]).

Le processus de collecte de données d'un réseau de surveillance dépend de son objectif. Les quatre objectifs suivants ont été identifiés[TDS⁺96, DH07] :

1. Déceler une maladie exotique ou inconnue
2. Hiérarchiser pour orienter la lutte
3. Suivre l'évolution de la maladie
4. Évaluer l'efficacité d'un plan de lutte

On parle de réseau d'épidémio-surveillance pour les objectifs 2 à 4 et de réseau d'épidémio-vigilance pour l'objectif 1. La collecte de données réalisée par un réseau de surveillance diffère selon l'objectif du réseau. Un réseau d'épidémio-vigilance par exemple, va demander un échantillonnage qui permette une forte sensibilité (détection d'un cas même si la prévalence (le nombre ou la proportion d'individus malades dans une population) est très faible) et ne nécessite pas un échantillonnage représentatif de la population (la prévalence observée dans l'échantillon ne doit pas nécessairement être représentative de la prévalence dans la population). Au contraire un réseau d'épidémio-surveillance mis en place pour évaluer l'efficacité d'un plan de lutte va privilégier un échantillonnage représentatif (qui permette d'estimer la prévalence dans la population) mais ne nécessite pas une détection de tous les cas. De nombreuses méthodes statistiques ont été développées pour estimer le type de collecte de

données à implémenter en fonction des objectifs de surveillance [TDS⁺96, DMS10].

Un réseau de surveillance est associé à un système d'information formalisé (et éventuellement informatisé) qui comprend la description précise de tout le parcours des informations, ainsi que les délais de transmission maximum des informations. Des diagrammes des flux de données et des informations résultantes sont disponibles où l'on retrouve le croisement des informations venant des différents composants du réseau de surveillance se superposant. Des points de contrôle des informations sont définis à chaque étape, et ce jusqu'au niveau le plus bas possible. Dans l'idéal il existe des bases de données permettant saisie et suivi des données depuis internet.

Néanmoins, des réseaux informels de surveillance peuvent court-circuiter ces systèmes d'informations. C'est particulièrement le cas dans les pays en voie de développement pour les maladies à contrôle obligatoire où la déclaration d'un cas dans un troupeau peut entraîner un abattage du troupeau qui ne sera pas indemnisé à l'éleveur. Des enquêtes sur ce sujet ont, par exemple, été réalisées par l'équipe AGIRs sur le cas de la grippe aviaire hautement pathogène au Vietnam. Les interviews d'éleveurs de volailles ont montré que les informations circulent entre les éleveurs de manière à anticiper des mesures de prophylaxie. Cependant, les données ne sont pas transmises au réseau formel ou ne remontent pas les niveaux d'intervention du réseau de surveillance parce que les acteurs du réseau veulent se prémunir de mesures de prophylaxie réglementaires (obligatoires) qui sont trop coûteuses à leur échelle (abattage au niveau de l'éleveur, quarantaine au niveau du village, ...) [DF11]. Les processus de remontée des informations et les protocoles de collectes de données peuvent donc être dans la pratique moins clairs que le système d'information ne les décrit.

Dans la mesure où la transition d'information au sein de ces réseaux peut être difficile à modéliser, nous nous attacherons dans cette thèse au processus de collecte de données, indépendamment du réseau qui a permis de récolter ces données. Nous utiliserons le terme de système d'épidémio-surveillance ou système d'épidémio-vigilance suivant l'objectif visé².

Ce qui doit être remarqué au sujet de la composition des réseaux de surveillance est qu'il existe de fait un lien ténu entre le système de surveillance et le système de prophylaxie (aussi appelé système de contrôle). Si l'on regarde de plus près le réseau de surveillance qui implémente le système de surveillance, on se rend compte aisément que certains acteurs (notamment l'éleveur) font partie intégrante des deux systèmes. Ce point est crucial dans l'argumentation de notre thèse car, en généralisant, on observe que l'on est en présence d'un système d'observation qui à la fois doit nous permettre de fournir un diagnostic sur une situation clinique, et à la fois influe sur cette situation.

2. La communauté vétérinaire française travaillant sur les réseaux de surveillance utilise le terme de système de surveillance pour désigner des réseaux de réseaux de surveillance organisés à l'échelle régionale (sous entendue supranationale). Nous devons donc préciser que le terme système ne désigne pas ici ce type de réseau particulier mais uniquement le processus dynamique de collecte de données implémenté par un quelconque réseau de surveillance. Il correspond au terme de « système de monitoring et de surveillance » utilisé dans la communauté anglo-saxonne (MOSS pour Monitoring and Surveillance System) [HPH09]

Évaluation des systèmes de surveillance

On peut distinguer deux types de méthodes d'évaluation d'un système de surveillance existant. Les premières consistent à évaluer la structure et le fonctionnement du réseau de surveillance qui implémente le système de surveillance. Ces méthodes se présentent généralement sous la forme d'un questionnaire standardisé qui vise à évaluer (de manière quantitative) l'adéquation entre la population à surveiller et les moyens mis en oeuvre par le réseau. Le type d'information récolté va par exemple être la quantité de bétail mesuré en Unité Gros Bovins (UGB) ou le financement du réseau en dollars américains pondéré par un indicateur du pouvoir d'achat en dollars américains [BAD⁺99]. Les secondes méthodes consistent à évaluer le système de surveillance en fonction des données qu'il récolte, en croisant par exemple les observations faites par le réseau avec celles réalisées sur la même population par un réseau différent [VGB11].

À ces méthodes il faut ajouter les méthodes d'évaluation de systèmes de surveillance théoriques qui utilisent la modélisation et la simulation. Le chapitre suivant présente un exemple de modélisation de réseaux de surveillance (cf. section 2.5.5).

La recherche sur l'évaluation des réseaux de surveillance est un enjeu particulièrement important dans les pays du sud exportateurs d'animaux vivants. Les règles du commerce international leur imposent de satisfaire des critères quantitatifs très stricts énoncés par l'Office Internationale des Épizooties (OIE) pour pouvoir être reconnus indemnes d'une maladie contagieuse. Donnons comme exemple la peste bovine pour laquelle l'OIE demande une collecte ponctuelle annuelle : *La procédure d'échantillonnage doit permettre de détecter la peste bovine avec une probabilité de 95% si elle est présente dans au moins 1% des troupeaux (ou des unités d'échantillonnage qui auront été définies). La surveillance clinique doit être maintenue pendant trois années successives pour aboutir au statut de pays indemne de maladie* [DH07]. Les seules méthodes qui permettent de remplir ce type d'exigence aujourd'hui sont des méthodes d'échantillonnage qui demandent beaucoup de moyens que les pays du sud ne sont pas toujours en mesure d'investir dans la santé vétérinaire.

L'évaluation des systèmes de surveillance fait donc partie des questions de recherche de l'équipe AGIRs et cette thèse a ainsi été associée au projet de recherche « Recherches pour l'EValuation en Asie du sud est de la Surveillance de l'Influenza Aviaire (REVASIA) » [VDB⁺10, VGS⁺10].

1.3 Les modèles et leur utilisation

1.3.1 Les modèles

Nous utiliserons la définition de « modèle » donnée par Marvin Minsky [Min65], à savoir :

To an observer B , an object A^ is a model of an object A to the extent that B can use A^* to answer questions that interest him about A .*

Cette définition est à la fois générale et claire. Elle est souvent adoptée dans les thèses de doctorat en modélisation, et l'est presque systématiquement dans la communauté scientifique dans laquelle nous évoluons [Bou94, Dub04, Bom09]. Cette communauté se place ainsi dans une posture où, par définition, le modèle n'est pas simplement une

copie dégradée de l'objet qu'il représente, mais il est indissociable de la question à laquelle il doit permettre de répondre, et de « l'observateur » qui le construit. Cette définition nous intéresse particulièrement et il est nécessaire de passer un peu de temps dessus car elle est au cœur de cette thèse. Tout d'abord elle sous-entend que pour étudier, évaluer, discuter d'un modèle, il nous faut bien sûr le considérer dans sa relation à l'objet qu'il modélise, mais également dans sa relation à l'observateur qui le construit et l'utilise et à la question à laquelle il doit répondre. Nous appellerons le triplet « objet, modèle et observateur » la triade de Minsky.

Considérons tout d'abord « l'objet ». Il ne s'agit pas nécessairement d'une partie du monde physique. Il peut s'agir de n'importe quel type d'objet. Il peut par exemple être lui-même un modèle.

Considérons ensuite l'entité désignée par le terme « observateur ». Dans l'article initial de Minsky, le terme « observateur » est rapidement remplacé par le terme « Homme » et il désigne en fait toute créature (selon les termes de Minsky) cognitive, c'est à dire toute créature capable de réfléchir sur elle-même et sur le monde. Minsky pose initialement sa triade pour discuter de la capacité d'introspection de cette créature et de ses limites. L'entité « observateur » ne se contente donc pas « d'observer » l'objet, il peut interagir avec lui. Plus encore, il construit et utilise le modèle en fonction de la relation qu'il entretient avec l'objet. Nous évoquerons par la suite la relation d'interaction entre l'observateur et l'objet par le terme de « Relation à l'objet » sans préciser la nature de cette relation. Dans le cas le plus classique, l'observateur expérimente l'objet en le stimulant et en observant la réponse de l'objet.

Considérons enfin l'entité « modèle ». Il s'agit d'une entité qui se substitue à l'objet dans la réflexion de l'observateur. Il peut s'agir par exemple d'un ensemble d'équations différentielles dont les variables peuvent être mises en relation avec des caractéristiques de l'objet observé. L'observateur va effectuer des opérations sur ces équations pour tirer des conclusions sur ces équations puis, par analogie, sur l'objet observé (état d'équilibre, comportement asymptotique etc.). De la même manière que pour la relation à l'objet, nous ne rentrerons pas dans les détails et parlerons pour l'instant de « relation au modèle ». Dans le cas de modèles de systèmes dynamiques, il peut s'agir de la réalisation de simulations du modèle qui consiste donc à réaliser un ensemble d'expériences sur l'entité modèle. La figure 1.3 illustre les entités qui composent la triade de Minsky et les relations qu'elles entretiennent entre elles. Dans toute la suite de la thèse, le signe « * » sera apposé à la suite d'un objet O pour désigner le modèle de Minsky O^* de cet objet.

La place de l'informatique

Les modèles mathématiques précèdent les sciences de l'informatique. Un modèle mathématique est élaboré en écrivant des équations qui décrivent par exemple des relations entre des variables d'intérêt sur l'objet. Ces équations permettent de formuler des questions dans un langage mathématique et de trouver des réponses en utilisant des règles mathématiques. On parle de résultats analytiques. Aujourd'hui, depuis l'essor des sciences de l'informatique, les capacités de calcul à notre disposition sont suffisamment grandes pour que l'on puisse utiliser des modèles que l'on ne peut résoudre analytiquement. C'est-à-dire que les équations qui permettraient de formuler ces modèles sont tellement complexes que les mathématiciens peuvent éven-

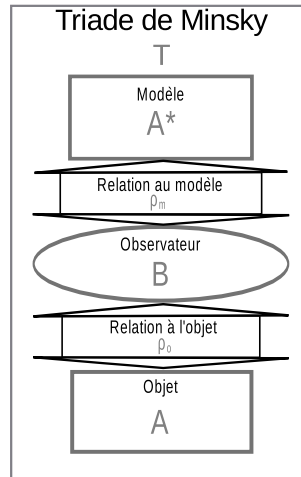


FIGURE 1.3: La triade de Minsky : Un *modèle* A^* de A est une entité qui permet à un *observateur* B de répondre à une question sur l'*objet* A . On note ρ_o et ρ_m les relations que l'observateur entretient avec respectivement l'objet et le modèle.

tuellement en extraire des propriétés intéressantes mais ne peuvent pas calculer de fonction analytique donnant l'état du modèle en fonction du temps. Il est nécessaire de calculer de manière itérative l'évolution de l'état du modèle à partir d'un état initial. Les utilisateurs de tels modèles utilisent alors des simulateurs qui permettent de réaliser des expériences sur ces modèles, de les simuler. On parle parfois d'expériences *in-silico* pour parler des simulations (il s'agit bien d'expériences réalisées sur un ordinateur). Notons cependant que cela reste une expérience sur le modèle de Minsky et pas sur l'objet de Minsky. La construction de ces simulateurs, leur spécification, ainsi que leur implémentation informatique font l'objet de nombreuses recherches et mobilisent de nombreuses personnes qui travaillent uniquement au niveau très abstrait et générique de la formalisation de processus dynamiques complexes et de la simulation de ces processus dynamiques sur des calculateurs [Tro02]. La Society for Computer Simulation (SCS) créée en 1952 est représentative d'une communauté de chercheurs qui développent des outils de simulation dédiés à des domaines d'application aussi différents que l'épidémiologie, la défense, les jeux vidéos ou la gestion des écosystèmes. La figure 1.4 positionne la recherche informatique en M&S dans le cadre de la triade de Minsky.

1.3.2 L'utilisation des modèles

Expériences, modèles et théories

Nous distinguerons les sciences expérimentales, qui s'intéressent au monde biophysique (comme la physique ou la biologie par exemple), des sciences abstraites qui ne cherchent pas à accumuler des connaissances sur le monde biophysique (comme les mathématiques). La démarche scientifique classique des sciences expérimentales est dite « inductive », elle consiste à généraliser des résultats expérimentaux de manière synthétique. Cette synthèse est une théorie si l'on pense qu'elle nous permet de prédire le résultat de nouvelles expériences à réaliser. Une théorie est en quelque

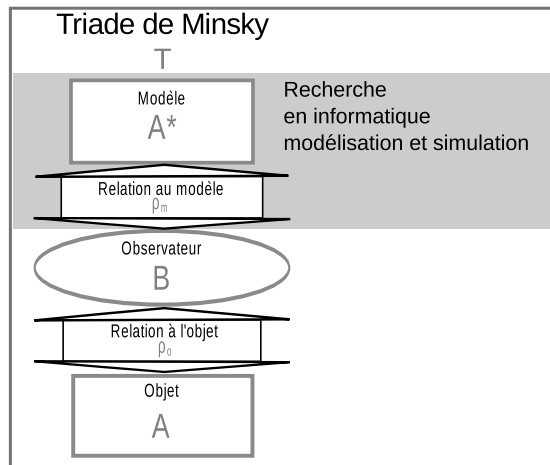


FIGURE 1.4: La recherche en informatique appliqué à la M&S dans la triade de Minsky : La recherche en informatique en M&S développe des méthodes et des outils pour permettre la simulation de modèles.

sorte une hypothèse sur les lois du monde qui a toujours été corroborée par l'expérience. Une théorie est considérée comme valable jusqu'à ce qu'elle soit réfutée par une expérience du monde biophysique qui ne la respecte pas. La figure 1.5 présente le cycle d'induction puis de déduction qui consiste à imaginer une théorie et à l'utiliser.

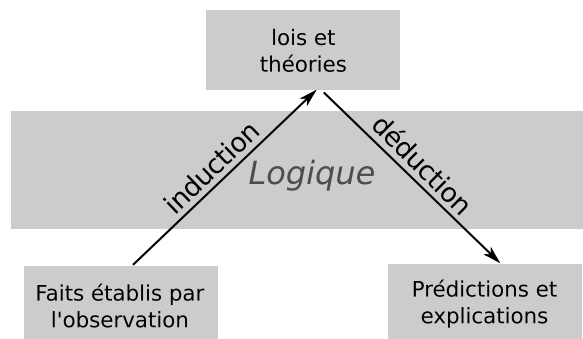


FIGURE 1.5: Démarche inductive, d'après [Bom09], [Cha82].

Une théorie s'énonce sous la forme d'un ensemble de principes ou de lois que le monde biophysique est supposé respecter. Pour prendre l'exemple de la physique, la théorie newtonienne avait initialement postulé un ensemble de trois principes qui régissaient la statique et la dynamique des corps : le principe fondamental de la statique, le principe fondamental de la dynamique et le principe des actions réciproques. Ces principes peuvent être exprimés sous la forme d'équations mathématiques. Ces équations, ainsi que les règles des mathématiques, constituent un modèle du monde. C'est-à-dire que les physiciens vont travailler sur ces équations pour répondre à des questions sur le monde. Mais la théorie de Newton n'est pas par essence ces équations. La théorie de Newton est que la dynamique et la statique des corps physiques

respectent les principes décrits par ces équations. Il s'agit d'une nuance importante. La confusion entre modèle et théorie vient certainement du fait que les lois ou principes des sciences physiques s'expriment de nos jours essentiellement sous forme mathématique et donc les théories qui se rapportent à ces domaines peuvent souvent être énoncées de la manière suivante : la théorie est que le monde physique se comporte de la même manière que ce modèle mathématique. Cette distinction est bien plus claire dans les sciences de la vie par exemple où l'énoncé de théories est plus difficile à mettre en équation (la théorie sur le vivant étant la théorie de l'évolution). Les biologistes prennent alors comme modèle une espèce, ou un patho-système dans le cas de l'épidémiologie, qui va leur permettre de répondre à des questions sur le vivant en général ou au moins sur des espèces proches de leur modèle. Les théories sur la morphogénèse par exemple ont été éprouvées sur le modèle de la mouche drosophile. Pour en revenir à la physique et à la manipulation de modèles mathématiques, la théorie de la relativité restreinte par exemple postule que la théorie de Newton n'est correcte qu'à la condition d'ajouter une nouvelle règle au monde. Cette nouvelle règle énonce comme une loi le résultat expérimental qui constate que la vitesse de la lumière ne dépend pas de la vitesse de l'observateur qui la mesure, et ce quel que soit la vitesse de l'observateur. En ajoutant cette règle au modèle mathématique décrivant les règles de la théorie de Newton, on obtient un nouveau modèle. Le fait que ce modèle nous permette de prédire le déplacement des étoiles et des électrons propulsés à grandes vitesses dans des accélérateurs de particules a pour conséquence que la théorie aujourd'hui acceptée est la théorie de la relativité restreinte. Jusqu'à ce que l'on observe un corps physique se déplaçant à une vitesse supérieure à celle de la lumière.

Notons que depuis quelques dizaines d'années, notamment avec l'apparition des ordinateurs, la démarche inductive que nous venons de présenter s'étend aux sciences abstraites telles que définies plus hauts. Le mathématicien Kurt Gödel avait ouvert une brèche en 1930 en démontrant que l'arithmétique, un des fondements des mathématiques, était incomplète, c'est-à-dire que certaines lois de l'arithmétique qui étaient vraies ne pouvaient être démontrées au sein de l'arithmétique [NNGG89]. Cela signifie que la méthode inductive, sous la forme de réalisation d'expériences, d'élaboration de théories et éventuellement d'utilisation de modèles serait nécessaire au sein même des mathématiques. Les mathématiciens parlent pour l'instant de conjecture pour désigner une assertion qui n'a jamais été réfutée mais qui n'est pas démontrée. En science de l'information, la théorie du chaos déterministe est un bon exemple. Une sélection de systèmes mathématiques simples sont utilisés comme modèles pour étudier la théorie du chaos qui dit de manière tout à fait générale que même un système fermé dont le comportement peut s'exprimer de manière simple et complète et qui possède un petit nombre de degrés de liberté, peut avoir un comportement totalement imprévisible à long terme [Lor63]. Ce caractère imprévisible est dû au fait que le système est très sensible à ses conditions initiales. On pourrait penser que la caractéristique « chaotique » ne pourrait s'appliquer qu'à des systèmes physiques dont on ne pourrait mesurer les conditions initiales de manière parfaite. Malheureusement il en va de même pour les systèmes mathématiques. Car on ne peut déterminer les trajectoires de ces systèmes que de façon itérative, à la main ou avec un ordinateur par exemple, et cela requiert de choisir une précision pour la mesure de l'état initial.

Les différentes utilisations des modèles

Nous distinguons trois objectifs d'utilisation de modèles en confrontation avec les observations réalisées sur un objet : la prédiction, l'explication, et la synthèse ou l'extrapolation de données. Dans un objectif de prédiction, le modèle de l'objet est utilisé pour identifier des scénarios possibles sur l'état futur, ou la trajectoire des états futurs de l'objet modélisé. Dans un objectif d'explication, le modèle est utilisé pour comprendre comment des phénomènes observés à une échelle donnée peuvent ou non être expliqués par un ensemble de principes ou de lois, éventuellement définies à des échelles plus fines. Dans un objectif d'extrapolation ou de synthèse de données, le modèle est utilisé pour représenter de manière synthétique un ensemble de données ou une tendance observée dans ces données. La figure 1.6 résume comment la modélisation est utilisée dans les différents domaines de l'épidémiologie abordés dans la thèse.

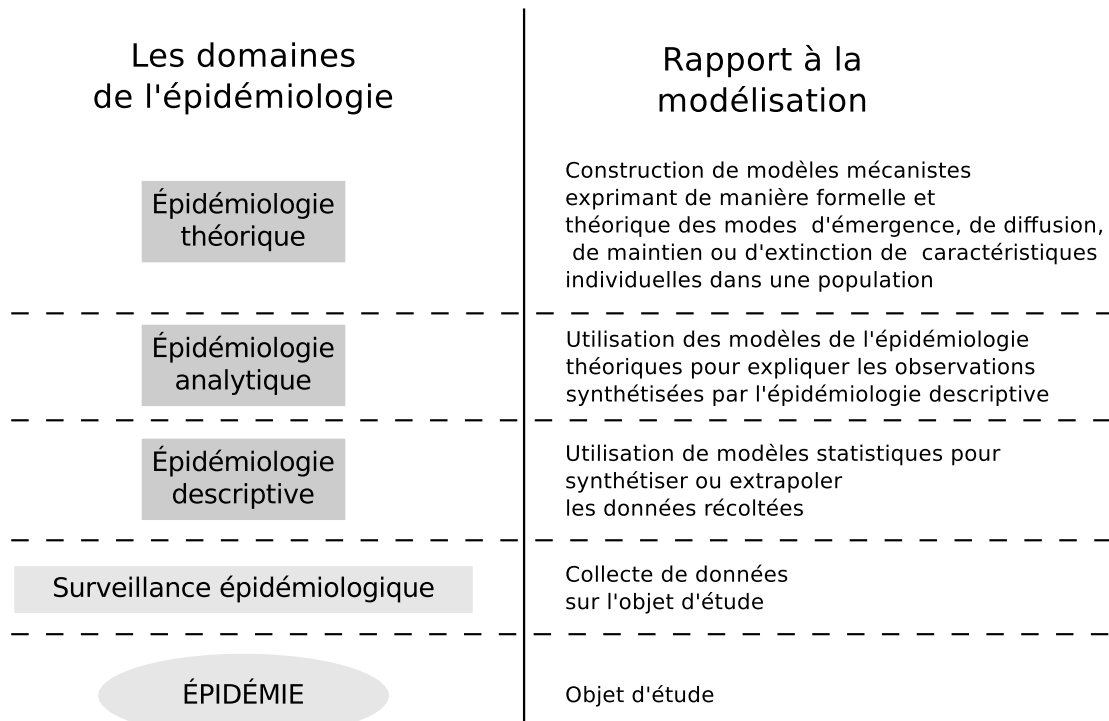


FIGURE 1.6: Domaines de l'épidémiologie et modèles

Un exemple en épidémiologie animale

Comme exemple en épidémiologie animale, nous présentons dans ce paragraphe une étude systématique du rapport entre les demandes des décideurs et les réponses apportées par les travaux de modélisation [SST11]. Cette étude a été rendue possible par l'élaboration d'une base de données de documents techniques qui porte sur l'utilisation des modèles en santé animale. L'autorité européenne de sécurité des aliments (EFSA pour European Food Safety Authority) dispose d'un groupe scientifique sur la santé et le bien-être des animaux (AHAW pour Animal Health and

Animal Welfare) qui a produit une série de documents qui résument les attentes des acteurs de terrain en terme de modélisation, et les modèles disponibles qui peuvent y répondre [Sin10]. Ces documents présentent une structure standardisée qui permet une analyse systématique. L'analyse a permis d'identifier des catégories de modèles et des types de questions adressées à ces modèles. En mettant en relation les types de questions et les catégories de modèles, les auteurs observent une forte corrélation. Les modèles entrant dans la catégorie que l'on pourrait décrire comme « les modèles d'inférence statistique » sont par exemple exclusivement utilisés pour répondre à des questions du type « caractérisation de test de diagnostic » (déterminer la sensibilité ou spécificité d'un test de sérologie par exemple). Ils sont utilisés pour synthétiser des données issues de test en laboratoire ou de la littérature. Les modèles qui entrent dans une catégorie « les modèles de systèmes dynamiques » sont par contre majoritairement utilisés pour des questions de type « apport de connaissance sur les processus de transmission des pathogènes ». Ils notent cependant des exceptions et émettent également l'hypothèse d'effets de mode qui dédient un type de modèle à une application particulière. De manière transversale à ces catégories de questions et de modèles, les auteurs ont définis trois types d'objectifs de modélisation, à savoir, description, prédiction ou compréhension. Ils notent un glissement fréquent d'un objectif de prédiction évoqué dans la question posée au modèle, vers un objectif de compréhension dans la description du modèle (l'objectif du modèle est identifié comme tel par les modélisateurs). Ce glissement est principalement attribué au manque de connaissances sur le système qui interdit toute prédiction fiable. Les auteurs soulignent que les résultats du travail de modélisation apportent néanmoins une aide aux décideurs en rationalisant la problématique de décision. Cette aide très informelle est malheureusement difficile à évaluer quantitativement. Cela eut été éventuellement possible si les documents avaient tracés les décisions qui ont été effectivement prises sur les sujets concernés mais ce n'est malheureusement pas le cas et les auteurs le déplorent [SST11].

Cette étude ne permet pas un exposé exhaustif de l'utilisation des modèles en épidémiologie vétérinaire, mais elle montre que la question de l'utilisation des modèles commence à se poser dans cette communauté et qu'il est possible d'étudier son impact sur les avancées en épidémiologie. L'effort d'analyse réalisé par le panel AHAW de l'EFSA est encourageant. Le fait que ce type d'initiative se concrétise nous incite en effet à penser que la dernière section de « décision finalement prise » qui manque aux documents devrait voir le jour dans de prochaines entreprises similaires à venir.

1.4 Problématique

Nous nous intéressons plus particulièrement dans cette thèse à des systèmes sur lesquels on ne peut pas réaliser d'expérience. Pour commencer, prenons l'exemple des systèmes caractérisés par une grande échelle spatiale qui nous interdit d'imposer une entrée au système. Il est possible d'utiliser des modèles pour prédire les comportements de ce type de systèmes à court ou moyen terme si les lois qui les régissent sont suffisamment stables. Les modèles climatiques, par exemple, fonctionnent car les lois de la physique macroscopique sont stationnaires à l'échelle temporelle considérée et les modèles peuvent être validés d'une journée sur l'autre ou d'une année sur l'autre. Dans le cas des systèmes sociologiques ou écologiques, les lois considérées

sont plus nombreuses et moins faciles à identifier qu'en physique macroscopique. Tout d'abord, elles sont individu-dépendantes, mais en plus elles sont sujettes à des changements que l'on ne maîtrise pas, comme l'apparition d'une nouvelle espèce dont on ne peut prédire la nature ou le comportement. Ainsi des écosystèmes structurés, stables en apparence, peuvent basculer très rapidement vers de nouvelles structures complètement différentes qui sont composées d'éléments différents et régies par des lois différentes. Les moments mêmes auxquels ont lieu ces basculements sont particulièrement difficiles à prédire [EU09].

De plus, la plupart des sciences étudiant ce type de phénomènes s'accordent aujourd'hui sur l'importance de considérer le système planétaire de façon holistique. Citons le courant de pensée « One Health » pour l'épidémiologie, ou l'approche écosystémique des pêches en halieutique. Cette interdépendance de nombreux phénomènes au niveau planétaire, fait qu'une situation passée ne réapparaîtra jamais à l'identique. De plus, l'importance que prend l'activité humaine dans l'évolution de l'ensemble des écosystèmes les rend difficile à observer en restant observateur passif complètement objectif.

Dans ces conditions, les études expérimentales classiques telles qu'on peut les trouver en sciences physiques, ou en biologie, sont proscrites. Dans les sciences expérimentales classiques, la méthode hypothéticodéductive est utilisée pour accumuler des connaissances. Une hypothèse est formulée puis testée par la suite par un plan d'expériences rigoureux qui permet soit de la réfuter, soit de définir son cadre de validité. Cela n'est plus possible lorsque l'expérience n'est plus reproductible. Bien sûr, des mécanismes généraux peuvent être identifiés, qui vont se répéter.

Beaucoup de mécanismes qui entrent en jeu dans la propagation d'une maladie par exemple, sont identifiés et vont se retrouver d'une épidémie à une autre. Néanmoins, la connaissance qualitative des mécanismes en jeu ne suffit pas à prédire l'évolution du système. Le jeu des interactions est tel qu'une connaissance précise du poids relatif de chacun de ces mécanismes est nécessaire. Or cette connaissance ne sera possible qu'une fois l'épidémie passée. L'exemple d'une épidémie est particulièrement intéressant, car en plus de nous fournir un cas de système pour lequel l'expérience n'est pas reproductible (on ne va pas volontairement provoquer une épidémie dans un pays pour étudier sa propagation), il nous fournit un cas où l'on ne peut même pas se permettre de rester un observateur passif (on ne va pas non plus laisser une épidémie se propager sans rien faire). En conséquence, en plus de ne disposer que d'une seule expérience, l'observateur fait partie intégrante de l'expérience puisqu'il ne peut rester passif. L'objectif de la section 1.2.2 portant sur les réseaux de surveillance épidémiologique était de souligner ce point. La trajectoire de l'épidémie va dépendre en partie du choix du réseau de surveillance utilisé pour la décrire et donc pour l'analyser.

En fait, il est non seulement difficile de prédire l'évolution de ce type de systèmes, mais il est également difficile de prédire l'impact qu'auront nos actions sur le système. Dans ces conditions, on peut questionner la pertinence des décisions que l'on prend. Aujourd'hui, la théorie de la décision permet de formaliser, et donc de rationaliser, la prise de décision dans l'incertain. Mais cela demande a minima une qualification ordinale du degré de possibilité que l'on attribue aux différentes conséquences que peut avoir une action.

On comprend que la construction de modèles qui puissent nous aider à prendre

des décisions sur de tels systèmes est ardue. Nous verrons dans notre état de l'art que de nombreuses méthodes existent pour palier aux difficultés rencontrées et qui naturellement permettent de prendre des décisions de manière « rationnelle ».

Néanmoins, nous verrons que ces méthodes n'abordent pas de front deux problèmes que nous soulignons ici. Tout d'abord, la non reproductibilité de l'expérience entraîne une validation sur une seule série de données, une seule histoire passée. En conséquence il est très difficile d'éprouver une théorie ou une loi qu'on aurait imaginée au sujet du fonctionnement d'un écosystème. Ensuite, notre objet d'étude (l'épidémie dans notre exemple), est dépendant des observations dont nous avons besoin pour construire notre modèle (collectées par le système de surveillance). Rappelons la définition que donne Marvin Minsky d'un modèle et l'illustration que nous en avons faite figure 1.3. Dans la mesure où la généralisation d'un modèle est rendue impossible par la non reproductibilité de l'expérience, la thèse que nous défendons est qu'il nous est néanmoins possible d'évaluer quantitativement et de manière rationnelle la pertinence d'un modèle A^* en terme de sa capacité à fournir à l'observateur B une réponse « convenable » sur l'objet A. Le caractère « convenable » de cette réponse pourra être évalué au regard de la trajectoire que la nouvelle connaissance produite avec A^* permet à B d'imposer à A (on parlera d'asservissement de l'objet A). En fait, davantage que le modèle, ce qui peut être intéressant de pouvoir évaluer, (et donc par la suite de pouvoir optimiser) est le processus d'observation de l'objet couplé au processus de construction et d'utilisation de modèles pour répondre à des questions sur l'objet ou à des objectifs d'asservissement de l'objet.

Par ailleurs, l'explosion exponentielle des capacités de calcul à notre disposition a permis d'ouvrir de nouveaux champs d'investigation à la science. Des modèles sont construits, qui permettent de réaliser des expériences virtuelles (des simulations) et qui permettent de tester des stratégies d'actions dans des mondes virtuels qui doivent présenter les mêmes caractéristiques que les mondes réels.

Nous pensons donc qu'il est possible et utile grâce à la capacité de calcul dont nous disposons aujourd'hui d'ouvrir un nouveau champ d'investigation qui étudie précisément les interactions entre l'objet modélisé, le modèle utilisé et l'observateur de Minsky qui construit le modèle et l'utilise, incarnant ainsi une rétroaction du modèle sur l'objet. Notre proposition prend la forme d'un cadre méthodologique et technique qui doit nous permettre de nous poser des questions sur la triade de Minsky (cf. figure 1.3). Notons T une triade de Minsky. Notre cadre doit permettre de construire un modèle T^* de cette triade et donner la possibilité de simuler le modèle T^* . Comme le montre la figure 1.7, le modèle T^* sera composé de trois entités. Un modèle A_C^* de l'objet, à mettre en relation avec l'objet A, un modèle B^* de l'observateur à mettre en relation avec l'observateur B, et un modèle du modèle A^{**} à mettre en relation avec le modèle A_B^* qui apparaît dans la triade T. On construit ainsi une nouvelle triade dans laquelle T^* est un modèle de T pour un observateur (noté C). Notons que le modèle A_B^* est un modèle A pour l'observateur B et que le modèle A_C^* est un modèle de A pour l'observateur C. De la même manière, le modèle A^{**} est un modèle de A_C^* pour l'observateur B^* ($(A_C^*)_{B^*}$) et un modèle de A_B^* pour l'observateur C ($(A_B^*)_C$). Rappelons enfin que nous avons réifié les relations de l'observateur à l'objet (ρ_o) et de l'observateur au modèle (ρ_m). La relation ρ_o désigne l'observation et éventuellement les actions que l'observateur effectue sur l'objet et la relation ρ_m désigne l'utilisation que l'observateur fait du

modèle.

La nouvelle triade obtenue (T' dans la figure 1.7) constitue le cadre général que nous proposons dans cette thèse. L'utilisateur de ce cadre y apparaît en tant qu'observateur C . La relation ρ_{o_c} désigne la relation que l'utilisateur de notre cadre entretient avec la triade T qui l'intéresse. Il nous semble important de la décrire mais nous ne proposons pas de cadre formel pour le faire. De manière informelle, il s'agit d'expliciter l'analyse que l'on peut avoir de la triade T qui nous intéresse et les modification que l'on veut éventuellement y apporter. La relation ρ_{m_c} désigne la relation que l'utilisateur de notre cadre entretient avec le modèle T^* . Cette relation peut en revanche être formalisée (et même automatisée) sous la forme de plan d'expériences à réaliser sur le modèle T^* .

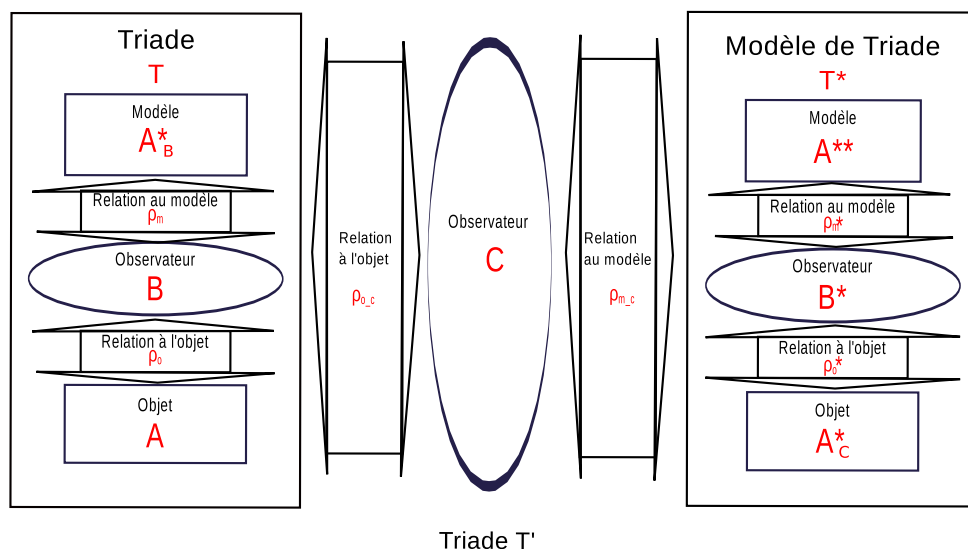


FIGURE 1.7: Cadre général : La triade T' .

Construire un modèle T^* de la triade T permet à un chercheur C d'adresser à T^* les questions qu'il se pose sur T . On obtient une nouvelle triade que l'on note T' (cf. figure 1.8).

La thèse que nous défendons est la suivante :

En modélisant et en simulant la triade de Minsky nous pouvons nous poser des questions sur les entités de la Triade de Minsky et sur leurs interactions. Cela nous permet de poser des questions sur l'asservissement de l'objet A par B en utilisant A^* .

1.4.1 Positionnement de la thèse en informatique

En faisant basculer la figure 1.7 on obtient une nouvelle triade dont l'objet à modéliser est une triade T . Nous avons noté T' cette nouvelle triade. La figure 1.8 présente ce basculement. L'effort de développement se portera sur la proposition d'un outil permettant de construire le modèle de triade (T^*) et de le simuler (ρ_{m_c}). Le premier objectif de ce manuscrit est de convaincre le lecteur que le cadre méthodologique et technique que nous proposons permet au chercheur en modélisation de

se poser de nouvelles questions sur son activité. Le second est de mettre à disposition ce cadre méthodologique et technique pour que les chercheurs thématiciens disposent d'un nouvel outil d'analyse de leurs recherches.

1.5 Plan de la thèse

Dans le chapitre suivant, nous présentons un état de l'art à la suite duquel nous revenons de manière plus précise sur la problématique de la thèse.

Dans le troisième chapitre, nous proposons un outil informatique générique pour construire et utiliser un modèle de triade de Minsky que l'on puisse expérimenter, c'est-à-dire que l'on puisse simuler.

Dans le quatrième chapitre, nous présentons l'application de ce cadre dans le contexte de la surveillance épidémiologique. Dans une première partie nous présentons la triade qui nous intéresse en surveillance épidémiologique. Nous la notons T_{uc} car elle porte sur le contrôle d'une épidémie (UC pour *under_control*). Dans une seconde partie nous présentons des questions de recherche sur cette triade. Dans une troisième partie nous présentons le modèle T_{uc}^* qui est un modèle de la triade présentée en première partie et les simulations à réaliser pour répondre aux questions sur T_{uc} .

Enfin, le dernier chapitre est réservé à une discussion sur l'apport de notre cadre en épidémiologie, sur les limites et les perspectives de l'outil que nous avons proposé, et sur l'apport et les limites du cadre que nous proposons de manière générale.

Dans la partie droite de la figure 1.8 nous précisons quelles parties du manuscrit sont dédiées à quelles parties de la triade T' .

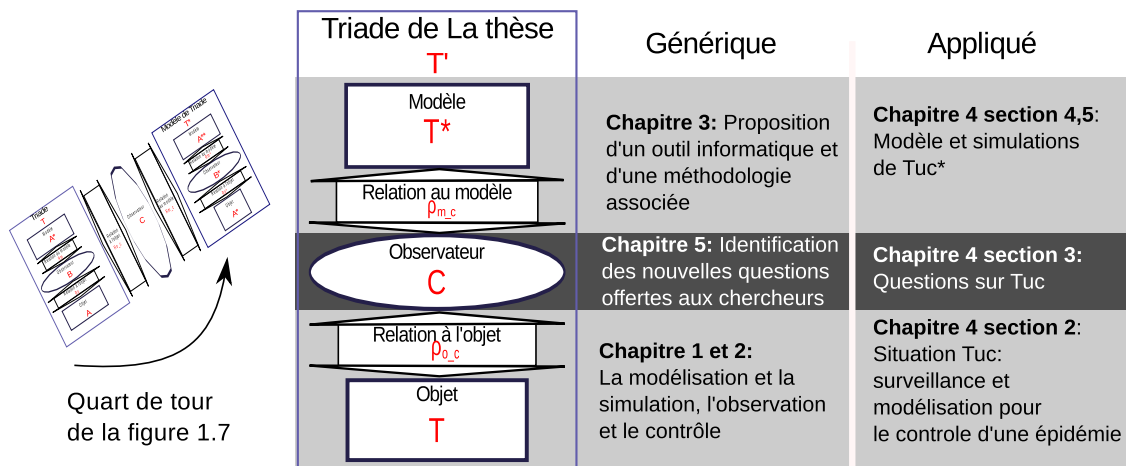


FIGURE 1.8: Organisation de la thèse autour de la triade T' : Cette thèse en informatique propose un outil qui permet de simuler la triade T^* . L'objectif du manuscrit est de montrer que de nouvelles questions sont posées. La partie droite de la figure positionne l'organisation de la thèse par rapport aux entités de la triade T' .

Chapitre 2

État de l'art

RÉSUMÉ

Notre posture épistémologique est celle de la systémique. Dans ce cadre, le modèle de triade de Minsky que nous proposons est un modèle de système dynamique. Il s'agit plus précisément de l'interaction entre un système cybernétique (l'observateur de Minsky) et son environnement (l'objet de la triade de Minsky). Les triades qui nous intéressent dans cette thèse sont celles dans lesquelles l'objet est un Socio-Éco-Système (SES) sur lequel on ne peut pas reproduire d'expériences. Dans ce chapitre, nous présentons un ensemble de travaux qui précisent la nature des entités des triades de Minsky qui nous intéressent (l'objet A , l'observateur B , le modèle A^* et les relations ρ_o et ρ_m) et qui doivent nous permettre de concevoir des modèles de ces triades. Le chapitre est conclu par la présentation de l'ensemble des outils fournis par la Théorie de la Modélisation et de la Simulation (TMS) qui vont nous permettre de spécifier et d'implémenter ces modèles.

2.1 Introduction

L'objectif de ce chapitre est de présenter l'état des connaissances actuelles qui vont nous permettre de construire un modèle du système triade que l'on puisse expérimenter. Suite à cette introduction sur la systémique, nous commencerons par présenter quelques caractéristiques des systèmes cibles qui nous intéressent de manière à pouvoir discuter plus aisément sur la nature de l'objet A des triades de Minsky qui nous intéressent. Nous présentons ensuite un ensemble de travaux sur la modélisation et la simulation. Certains de ces travaux en modélisation nous intéressent car ils permettent de construire les objets A_B^* des triades de Minsky qui nous intéressent. D'autres travaux sont présentés dans cette section, qui ne portent pas précisément sur l'objet A_B^* des triades de Minsky qui nous intéressent, mais qui vont nous aider à modéliser ces triades. Nous présenterons ensuite un ensemble de méthodes d'observation et de contrôles qui correspondent à la relation ρ_o que l'observateur de Minsky entretient avec l'objet. De la même manière, nous présentons un ensemble de travaux concernant différentes utilisations possibles des modèles de

systèmes cibles, représentant la relation ρ_m que l'observateur de Minsky entretient avec le modèle A_B^* et notamment la manière dont il expérimente A_B^* . Nous présenterons enfin un ensemble de méthodes existant pour formaliser la prise de décision et qui peuvent être utilisés pour la modélisation du processus de décision B. Nous conclurons en présentant les outils fournis par la théorie de la modélisation et de la simulation qui vont nous permettre de construire le modèle de système triade et de l'expérimenter.

2.1.1 Théorie systémique et cybernétique

La théorie systémique consiste à se détacher du réductionnisme en considérant certaines parties de la réalité comme des entités dont on observe le comportement. Elle se focalise sur l'étude des relations entre ces entités et leur environnement ou éventuellement sur l'étude des interactions entre des entités distinctes. La systémique fournit également un langage commun à des scientifiques de disciplines différentes pour décrire les propriétés de systèmes dynamiques de manière générale. Elle permet par exemple de définir de manière rigoureuse des caractéristiques générales de phénomènes dynamiques comme l'équilibre ou la stabilité [Wal77].

Lorsque le système est capable d'observer son environnement et lui-même, on parle de système cybernétique. Étymologiquement, cybernétique est dérivé d'un verbe grec signifiant « piloter ».

Avec la théorie systémique arrive une nouvelle épistémologie, dans le sens littéral d'une nouvelle science des connaissances. Dans une vision systémique du monde, les connaissances accumulées sur le monde doivent être rattachées à des systèmes qui ont une limite et une échelle spatio-temporelle identifiée et à l'environnement dans lequel on considère ces systèmes. Un système reçoit de la matière et des informations de son environnement et lui renvoie de la matière et des informations. On dit que les flux de matière et d'information reçus sont les entrées du système et que ceux renvoyés sont les sorties du système. On parle de comportement d'entrée/sortie pour décrire la relation observée entre les entrées et les sorties du système. La limite et l'échelle spatio-temporelles d'un système ne sont pas déterminées par la nature des phénomènes observés, mais un choix réalisé par l'observateur qui a accumulé les connaissances. Bien sûr, on peut tout de même dégager certaines règles pour la définition de ces limites. Par exemple on va généralement délimiter un système de telle manière que les interactions internes au système soient plus nombreuses que les interactions entre le système et son environnement.

L'essor de la cybernétique dans les années 50 [Wei48], va de paire avec celui des sciences cognitives telles qu'elles sont décrites au sens large par Varela qui ne donne pas de définition précise mais en fait un exposé qui reste d'actualité dans [Var88]. Pour construire un modèle de la triade de Minsky que l'on puisse expérimenter, nous nous plaçons dans la posture épistémologique de la systémique. Ce faisant, nous considérons l'observateur de Minsky comme un système cybernétique (nous emploierons également le terme de « système cognitif » de manière équivalente) qui modélise le monde (ou une partie du monde) dans lequel il évolue. En cela, nos recherches en informatique appliquées à la modélisation et à la simulation s'apparentent aux sciences cognitives. Varela distingue trois grands courants de pensées dans les sciences cognitives, le courant symbolique, le courant connexionniste et le

courant enactionniste [Var88]. Le courant symbolique considère la cognition comme la construction et l'utilisation d'une représentation symbolique du monde. Le courant connexionniste s'oppose au courant symbolique et considère la cognition comme un ensemble de fonctionnalités qui émergent dans un système cybernétique pour lui permettre d'interagir avec son environnement. Le courant enactionniste, défini par Varela, consiste à relier la structure d'un système cognitif à son histoire vécue et expliquer ainsi l'émergence de fonctionnalités cognitives par une interaction entre l'environnement et le système cognitif (cf. [Var88, chapitre 4] et [VTR93]). Nous pensons que le courant enactionniste est le plus à même de représenter le raisonnement à l'échelle d'un être vivant. Cependant, rappelons que notre thèse porte sur l'utilisation des modèles pour l'étude ou la gestion des systèmes épidémiologiques, sociologiques et écologiques. Dans ce cadre, nous ne nous intéressons pas à un système cognitif du type d'un être vivant, mais à un système cognitif résultant d'une démarche scientifique rationnelle, qui utilise des modèles mathématiques et des méthodes d'analyse pour déterminer le modèle à utiliser et l'utilisation à en faire pour étudier ou contrôler le système observé. En conséquence, nos recherches s'apparentent au courant de pensée symbolique des sciences cognitives qui considèrent la cognition comme la construction et l'utilisation d'une représentation symbolique du monde.

Dans ce cadre, une limite, ou au moins un paradoxe, a été identifié dans le système de construction de connaissances absolues sur le monde d'un système cognitif. Nous allons résumer ce paradoxe dans le paragraphe suivant, en reprenant l'énoncé qu'en fait Robert Vallée dans [Val91].

2.1.2 La cognition et le paradoxe de « l'observation - action »

Tout système capable de cognition, un chercheur par exemple, est un système cybernétique. Notons S ce système. En tant que système, il est délimité dans le temps et l'espace et on peut désigner comme étant son environnement, l'ensemble de l'univers qui ne fait pas partie de S . Robert Vallée note F l'environnement de S et U la somme de S et F qui constitue l'ensemble de l'univers. Pour un système cybernétique, il intervient nécessairement une chaîne informationnelle, (elle accumule de la connaissance et Robert Vallée l'appelle l'étape épistémologique) que l'on décrit par un *opérateur d'observation* O , et une chaîne décisionnelle (étape praxéologique) qui fait intervenir un *opérateur de décision* D . Mais la distinction du fonctionnement de S en deux chaînes indépendantes n'est qu'un point de vue arbitraire et la seule certitude est que l'on a en fait qu'une seule chaîne que Robert Vallée appelle la chaîne pragmatique et qui implique l'*opérateur pragmatique* $P = DO$. Le paradoxe de « l'observation - action » naît de ce caractère indissociable des opérateurs D et O . Robert Vallée parle de « clôture épistémo-praxéologique » du système sur lui-même et sur son environnement (cf. figure 2.1). Citons le dernier paragraphe du chapitre de Robert Vallée sur ce paradoxe [Val91] :

L'univers U est perçu et agi par le système S qui est, en fait, une partie de U . Chaque individu connaît le monde et agit sur lui à travers cette partie du monde qu'il constitue. Le monde est paradoxalement perçu et agi par des portions de lui-même et non pas par des entités qui lui

seraient étrangères. Les systèmes qui le connaissent et le transforment, ainsi que leurs organes de perception, de décision et d'effection, sont faits des mêmes matériaux que lui. Il y a peut-être là le point de départ d'une harmonie préétablie d'un nouveau genre qui aiderait ou limiterait, de façon intrinsèque, l'intelligibilité et la transformabilité du monde.

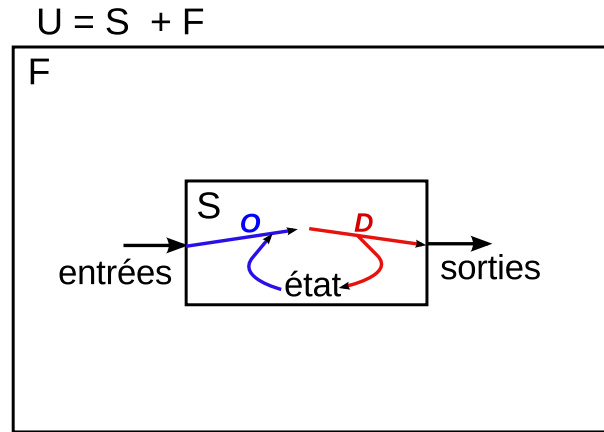


FIGURE 2.1: La Cognition et le paradoxe de l'observation - action (d'après R. Vallée [Val91])

Le système S et son environnement F constituent l'univers U . Le comportement du système cognitif S peut être vu comme la réunion d'un opérateur d'observation O et d'un opérateur de décision D . S fait partie intégrante de U et il ne peut dissocier entièrement ses observations de ses actions. En conséquence, sa vision de l'univers U n'est pas objective.

Le paradoxe de l'observation-action énoncé par Robert Vallée est à notre avis lié aux problèmes que nous rencontrons pour modéliser des systèmes sur lesquels une expérience n'est pas reproductible. En effet, si une expérience n'est pas reproductible, il est impossible de mesurer l'impact des actions de l'observateur sur le déroulement de l'expérience car on ne pourra pas renouveler l'expérience en changeant uniquement ces actions. En ce sens, l'observateur de Minsky est un système cognitif (S) qui fait partie de l'univers qu'il observe.

2.2 A : Les systèmes épidémiologiques, sociologiques et écologiques

2.2.1 Les systèmes complexes

Les systèmes tels qu'ils sont définis par la systémique ne désignent pas nécessairement des objets du monde physique. Ils peuvent également désigner des objets abstraits comme des structures mathématiques ou des logiciels, ou des parties de logiciels informatiques. Certains systèmes sont simples et prévisibles. On connaît la chaîne causale qui relie les entrées et les sorties du système et il est aisé de les contrôler. C'est-à-dire que l'on peut déduire le signal d'entrée que l'on doit imposer au système pour obtenir le signal de sortie que l'on désire. À l'opposé, certains

systèmes semblent avoir un comportement complètement chaotique et il semble impossible ou très difficile de prédire leurs comportements ou d'asservir leurs sorties en imposant une boucle de rétroaction de la sortie du système sur son entrée. Les systèmes qui nous intéressent ici (épidémiologiques, sociologiques et écologiques) appartiennent davantage à cette deuxième catégorie de systèmes.

Systèmes complexes et systèmes chaotiques

Si l'on suppose que l'on connaît les règles qui gouvernent le système, et que ces règles sont déterministes, deux raisons ont été identifiées qui le rendent difficile à prédire ou à contrôler. Nous dirons par la suite que le système est soit complexe, soit chaotique (Il peut également être complexe et chaotique). Présentons brièvement la théorie de la complexité et la théorie du chaos déterministe pour mieux définir le caractère complexe ou chaotique d'un système.

Systèmes complexes. La complexité d'un problème peut se mesurer comme le rapport entre le nombre d'opérations à effectuer pour résoudre le problème et la taille de l'énoncé du problème [Car08]. On parle de complexité de Kolmogorov ou de complexité algorithmique. Un exemple classique de problème complexe est le problème du voyageur de commerce. Il peut s'énoncer de façon informelle de la manière suivante : Un voyageur de commerce doit visiter l'ensemble des villes situées dans une région dont il connaît le réseau routier. Quel est le trajet le plus court qui lui permettra de passer par l'ensemble des villes en dépensant le moins d'essence. De manière plus formelle, le problème est le suivant : Soit un graphe valué G , quel est le plus court chemin passant par tous les nœuds de G ? Notons maintenant P_G le problème du voyageur de commerce pour le graphe G et X_G la taille de l'énoncé du problème P_G (X_G dépend du nombre de nœuds dans G). On peut montrer qu'il est impossible de trouver une fonction polynomiale $f : \mathbb{N} \rightarrow \mathbb{N}$, telle que le nombre d'opérations à effectuer pour résoudre le problème P_G soit toujours inférieur à $f(X_G)$ quel que soit G . On dit que le problème n'est pas résoluble en temps polynomial (il est NP complet). C'est en ce sens que le problème du voyageur de commerce est considéré comme complexe.

En revanche, il n'existe pas aujourd'hui de théorie des systèmes complexes unifiée qui permettrait de mesurer de manière quantitative la complexité d'un système de la même manière que la complexité de Kolmogorov permet de quantifier la complexité d'un problème. La complexité d'un système se définit davantage de manière informelle par le fait que le système est composé de composants distincts présentant de nombreuses interactions et dont la connaissance du comportement des sous-systèmes ne permet pas de prédire « de manière simple » le comportement du système dans sa globalité. Nous conserverons cette définition informelle de système complexe. Notons toutefois que cette définition peut être précisée par des critères quantitatifs. On peut par exemple qualifier de complexe un système composé d'un grand nombre de composants en interactions et structuré de telle sorte que le nombre de calculs à effectuer pour déterminer les changements d'états de l'ensemble du système soit une fonction combinatoire du nombre de composants. La communauté française travaillant sur ce type de systèmes se regroupe tous les ans lors des « Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels » de Rochebrune.

Systèmes chaotiques. Nous avons évoqué la théorie du chaos dans l'exposé de notre problématique. Elle postule qu'il existe des systèmes dynamiques dont le comportement est déterministe et dont, quel que soit le niveau de précision auquel on connaît les conditions initiales, il y aura toujours une date au delà de laquelle on ne pourra garantir une erreur maximale donnée de prédiction. Nous qualifierons donc de chaotique un système qui n'est pas nécessairement composé d'un grand nombre de composants et qui ne présente pas forcément de nombreuses interactions mais dont le comportement se définit de manière itérative et dont l'état et le comportement à long terme sont sensibles aux conditions initiales.

2.2.2 Résilience des systèmes sociologiques et écologiques

Certains systèmes perdurent dans le temps. Cette persistance (de structure et ou de comportement) les rend facile à isoler du reste du monde pour un observateur (qui peut lui-même faire partie du système). Cette persistance a lieu malgré de constantes perturbations venant de l'extérieur du système. On dit qu'un système est résilient lorsqu'il est capable de retrouver son état ou une partie de son état (sa structure par exemple) suite à une perturbation. Les exemples de systèmes résilients sont typiquement les êtres vivants ou les écosystèmes. Varela et Maturana ont ainsi défini le concept d'autopoïèse pour caractériser un organisme vivant comme un système cybernétique capable de rendre résilient sa propre organisation [VMU74]. À l'échelle des systèmes sociologiques et écologiques, certains chercheurs en écologie et en sciences sociales ont étendu le concept d'écosystème à celui de Socio-Éco-Système dans la mesure où la plupart des écosystèmes étudiés aujourd'hui sont anthropisés. Ces chercheurs proposent d'analyser les Socio-Éco-Système en les considérant comme des systèmes auto-organisés dotés d'une certaine résilience [Fol06].

2.2.3 Les Socio-Éco-Systèmes

Le concept de Socio-Éco-Système (SES) nous semble le mieux décrire le type de systèmes cibles qui nous intéressent dans cette thèse. Ces systèmes font intervenir des processus écologiques et des processus sociologiques. Les processus épidémiologiques peuvent être inclus pour certains parmi les processus écologiques (la co-évolution des pathogènes et de leurs hôtes par exemple) et pour d'autres parmi les processus sociologiques (les mouvements d'animaux vivants dans une filière de production animale par exemple). L'échelle considérée est celle d'une population humaine, elle s'étend à un grand nombre d'individus et sa granularité est l'humain. L'étude des SES a émergé dans le cadre de l'étude du réchauffement climatique dans le but de trouver des moyens d'y faire face ou de s'adapter [Gal91]. Le concept de SES est développé comme un cadre d'étude interdisciplinaire de plus en plus reconnu. Une formalisation en a été donnée par le prix nobel d'économie Elinor Ostrom [Ost09]. Nous n'utilisons pas le cadre proposé par Elinor Ostrom mais nous nous intéressons au même type de systèmes que nous désignerons donc sous le terme SES. Citons également le courant de pensée « *political ecology* » qui s'intéresse au même type de système mais l'analyse avec le point de vue des sciences politiques [GP94].

2.3 A_B^* : Les modèles

L'objectif de cette section est de présenter les modèles utilisés pour représenter le système cible dans les triades de Minsky qui nous intéressent. Il s'agit de modèles de systèmes dynamiques. Avant de présenter ce type de modèles, nous présentons les modèles disponibles pour représenter l'incertain car ils sont à la base de nombreux travaux de modélisation. Nous présentons également le paradigme multi-agent qui est aujourd'hui beaucoup utilisé pour l'étude des SES. Nous présentons enfin les travaux portant sur les modèles de modèles et notamment une étude réalisée dans le cadre de cette thèse qui nous semble bien illustrer les questions se posant au jourd'hui sur la modélisation des systèmes épidémiologiques.

2.3.1 Modèles de représentation de l'incertain

La communauté scientifique dispose aujourd'hui de deux cadres théoriques pour représenter l'incertain. On peut le représenter soit de manière quantitative en utilisant la théorie des probabilités, soit de manière qualitative avec la théorie des possibilités. La théorie des probabilités a été développée au XVII^e siècle et est suffisamment connue pour qu'aucun rappel ne soit nécessaire ici. La théorie des possibilités est beaucoup plus récente et nous vient des développements en logique floue et en modélisation du raisonnement [Zad99, DP01]. Cette théorie consiste à remplacer les probabilités qui, utilisent le caractère cardinal des réels pour mesurer le caractère « probable » d'une éventualité, par des possibilités qui utilisent le caractère ordinal des réels pour mesurer le caractère « possible » d'une éventualité.

Nous n'allons pas développer ici les multiples travaux développés au sujet de ces théories. Nous en parlons dans ce paragraphe précis parce que de nombreux modèles sont basés sur ces représentations de l'incertain. Certains modèles consistent exclusivement à construire une représentation du monde dans le cadre de ces théories. On peut citer en épidémiologie les modèles dits « d'arbres de scénarios » (par exemple [FAESL09]) ou de manière générale, les modèles d'analyse de risque probabilistes.

2.3.2 Les modèles de systèmes dynamiques

Ils sont aussi appelés modèles de simulation. Nous utiliserons le terme de Modèle de Système Dynamique (MSD) pour en parler car le terme « modèle de simulation » peut parfois être utilisé en opposition aux « modèles analytiques » qui peuvent être résolus sans simulation. Il s'agit de modèles mécanistes, c'est à dire qu'ils décrivent des processus qui traduisent des relations de cause à effet entre l'état du système et de ses entrées à un instant donné et la valeur de son état et de ses sorties à l'instant qui suit. Les MSD sont traditionnellement spécifiés sous la forme de systèmes d'équations différentielles qui décrivent l'évolution d'un ensemble de variables d'intérêt en fonction du temps. Nous avons choisi le modèle épidémiologique Susceptible, Infectieux, Susceptible (SIS) [AM79] comme exemple d'illustration. Il se présente sous la forme d'un système d'équations différentielles qui décrit la dynamique de deux variables $(S, I) \in \mathbb{R}^2$, respectivement la proportion d'individus susceptibles et infectieux dans la population (cf. équations 2.1, 2.2).

$$\frac{dS}{dt} = -\beta IS + \gamma I \quad (2.1)$$

$$\frac{dI}{dt} = \beta IS - \gamma I \quad (2.2)$$

Le paramètre γ correspond au taux de guérison moyen des individus de la population. Le paramètre β est moins facile à appréhender intuitivement. Ce paramètre peut s'exprimer sous la forme $\beta = -k \log(1 - c)$ où k est le nombre de contacts moyen entre deux individus par unité de temps et c est la probabilité qu'un contact entre un individu susceptible et individu infectieux entraîne une infection. Le modèle SIS a été construit sur cette base en supposant une population avec un mélange homogène qui permet d'utiliser une loi d'action de masse. Dans la majeure partie des cas, les paramètres k et c ne sont pas connus. Dans ces conditions, le paramètre β est estimé en utilisant des méthodes de calibration (cf. section 2.5.4).

Outre les équations différentielles, il existe de nombreux autres formalismes pour représenter des systèmes dynamiques. Automates à états finis, réseaux de Pétri, automates cellulaires, en sont quelques uns. La section 2.7 est dédiée à la formalisation des MSD.

La littérature sur les modèles épidémiologiques est abondante et il ne nous semble pas utile de la reprendre ici. Nous renvoyons le lecteur à l'ouvrage de M.J. Keeling et P. Rohani où il trouvera un état de l'art de la modélisation des maladies infectieuses chez les humains et les animaux particulièrement didactique [KR07]. La section 2.5.5 présente un exemple de MSD utilisé dans le cadre de la surveillance épidémiologique.

2.3.3 Le paradigme multi-agent

Le paradigme multi-agent est fréquemment utilisé pour étudier des systèmes complexes, particulièrement dans le cas des systèmes écologiques ou sociologiques [BP04]. Il s'agit d'une façon de voir un système complexe sous la forme d'un ensemble d'agents qui interagissent entre eux et éventuellement avec les objets d'un environnement qu'ils ont en commun et dans lequel ils évoluent [Fer95]. L'exemple d'une colonie d'insectes, souvent repris, semble une bonne illustration d'un système multi-agent. Le système « colonie de fourmis » a un comportement macroscopique que l'on peut décrire relativement facilement. Il est, par exemple, capable de se défendre et de s'approvisionner en énergie en trouvant de la nourriture à l'extérieur de la fourmilière et en la ramenant à l'intérieur de manière efficace. Le système « colonie de fourmis » est un système complexe et les agents qui le composent (les fourmis) s'adaptent de manière à assurer la survie du système « colonie de fourmis » dans son ensemble [Dro93]. C'est un cas emblématique où un système clairement identifiable par un observateur (la fourmilière) est défini au niveau d'agents autonomes qui la composent, et bien sûr par l'environnement dans lequel ces agents sont amenés à évoluer.

De la même manière ce paradigme permet de se poser des questions sur les sociétés humaines en aidant à faire le lien entre les règles que l'on connaît à des échelles individuelles et des comportements observés à l'échelle d'un système de plus grande échelle.

Les modèles à base d'agents semblent particulièrement adaptés à l'étude des systèmes sociologiques ou écologiques [TDZ08]. Ces modèles peuvent être les mêmes objets mathématiques que des modèles dit individus-centrés qui décrivent les dynamiques d'une population d'individus ou de particules par les règles de comportement des individus ou des particules. Néanmoins, ce changement de paradigme est apprécié car il permet de rapprocher les discours des acteurs du SES avec celui des modélisateurs et des chercheurs en sciences humaines et sociales et en écologie, notamment en permettant la définition d'individus/agents hétérogènes et éventuellement complexes [FMB10].

Des outils de formalisation de règles de fonctionnement individuelles issues des sciences humaines et sociales peuvent être aisément intégrés dans des systèmes multi-agents. Cela permet via ce paradigme de formuler des questions de recherches sur des phénomènes qui sont déterminés à plusieurs échelles différentes [BPT05, BPPTss], c'est à dire au niveau de la définition des règles individuelles des agents et au niveau de la structure de la population d'agents, éventuellement déterminée par l'agence-ment de l'environnement.

Nous n'utilisons pas le paradigme multi-agent dans le cadre de nos travaux mais il est utilisé en épidémiologie, notamment dans la communauté scientifique qui nous entoure [LPFL09, ADD08] et nous pensons qu'il serait intéressant dans le cadre de nos travaux.

2.3.4 Les modèles de modèles

En modélisation de systèmes dynamique, construire un modèle de modèle consiste à construire un modèle de système dynamique qui ait le même comportement qu'un autre modèle. Cela peut se faire par exemple dans le cas où un modèle est trop long à simuler et où on le modélise avec un autre modèle. Cette technique est utilisée pour changer d'échelles dans le cas où on dispose d'un modèle individu centré et où l'on veut simuler une grande population. On peut alors soit calibrer un modèle plus simple sur les résultats de simulations [SDD⁺04, DDS⁺04, BQMDew]. Soit, dans le cas de modèles mathématiques, on peut utiliser des méthodes mathématiques de simplification de modèles [Dur94, LD96]. Nous présentons en annexe une étude complète de l'utilisation de ce type de simplification de modèle pour le cas d'un modèle épidémiologique individu-centré représentant la propagation d'une maladie dans une population d'hôtes mobiles (cf. annexe E). Cette étude illustre bien le type de questions qui peut se poser au sujet des modèles épidémiologiques et nous donnons dans la section suivante les motivations de cette approche. Nous renvoyons le lecteur à l'annexe E pour une présentation des développements effectués et des résultats obtenus sur ces travaux.

Approximation par moments d'un modèle individu-centré de propagation de maladie dans une population d'hôtes mobiles

La structure du réseau de contacts au sein de la population d'hôtes joue un rôle majeur dans la dynamique de propagation d'une maladie infectieuse. Différentes structures de contacts vont entraîner différentes dynamiques d'infection [SR05, PJ08]. Or ces dynamiques ne peuvent pas être reproduites par des modèles de champs

moyen qui font l'hypothèse d'un taux de contact homogène et constant entre individus sains et infectieux [SM09]. En effet, un phénomène de regroupement des individus infectieux apparaît au cours de l'épidémie et en raison de ce phénomène, on observe un taux d'infection par individu infectieux, qui varie au cours de l'épidémie [FK99, KW99] et dont la dynamique dépend de la structure du réseau de contact [BBPSV05].

Pour cette raison, de plus en plus de modèles considèrent les maladies infectieuses comme la diffusion de maladies à travers un réseau de contacts entre les individus d'un groupe [KE05, Fra04]. La question se pose alors de savoir comment connaître ce réseau de contacts. Dans certains cas, il peut être obtenu en enquêtant les populations concernées, cela est fait en épidémiologie humaine (voir [ERE09] par exemple) et en épidémiologie animale à des échelles adaptées (voir les nombreux travaux sur la FMD au Royaume Uni [KWS⁺01] par exemple).

Le réseau de contact est souvent difficile à obtenir et pour cette raison, un des défis actuel de la modélisation en épidémiologie est d'arriver à reproduire au mieux les dynamiques d'infections à l'échelle des populations, en utilisant une information minimum sur la structure de contact entre les individus. La première étape est de déterminer les situations et les échelles pour lesquelles cette structure de contact joue un rôle important [Kee05, SR05, TM01]. Ensuite, dans les cas où cette structure de contact est primordiale mais où le réseau de contact n'est pas disponible, plusieurs options s'offrent au modélisateur. Dans le cas où certaines statistiques sur la structure du réseau sont disponibles, il est possible d'utiliser des modèles qui approximent l'impact du réseau sur la dynamique de propagation de la maladie [BBPSV05, HK10]. Malheureusement ces statistiques elles-mêmes ne sont pas toujours disponibles. C'est le cas en épidémiologie animale où les méthodes de construction de ces statistiques venant de l'écologie ne sont pas encore prêtes à fournir les statistiques nécessaires aux modèles épidémiologiques de manière fiable [PCS⁺09].

Dans ce cas précis, il est nécessaire de simuler les comportements individuels responsables des contacts entre individus en utilisant des modèles individus centrés qui simulent les comportements animaux [TGM⁺99, SDD⁺04, ADD08]. Les modélisateurs se confrontent alors aux problèmes usuels liés aux modèles individus-centrés qui sont la plupart du temps stochastiques, longs à simuler pour de grandes populations, et qui interdisent toute étude analytique. Dans l'étude présentée annexe E nous proposons d'utiliser la méthode d'approximation par moments [DL00, LMD03] pour construire un système d'équation-différentielles qui permette d'approximer ces IBM.

Cette méthode a montré son intérêt pour l'approximation de modèles individus-centrés de diffusion de maladie dans un graphe [TM01, FG01, Bol99]. Ici, nous l'appliquons directement sur un modèle individu centré qui simule le déplacement des individus. Une telle approximation doit nous permettre de travailler sur un modèle qui soit déterministe, mathématiquement analysable, rapide à simuler et qui prenne en compte la structuration de la population dans la dynamique de propagation de la maladie. Cela sans utiliser aucune statistique sur le réseau de contacts entre les individus mais en se basant sur la description des comportements individuels.

Nous prenons l'exemple de la propagation du virus hautement pathogène H5N1 dans un élevage de volailles. Cela nous permet de justifier d'une bonne homogénéité de l'environnement (un hangar par exemple) et de la population d'hôtes (une cohorte

de poulets), et de nous concentrer sur l'impact du déplacement des individus dans les dynamiques épidémiologiques.

L'annexe E a un format d'article scientifique et est rédigée en anglais. Dans une première partie, nous construisons le modèle individu centré en lui donnant une formalisation mathématique. Dans une deuxième partie nous partons de cette formalisation mathématique pour construire le modèle approximé en utilisant une approximation par moments. Dans une troisième partie nous rappelons le modèle épidémiologique de champs moyen qui n'est autre qu'une approximation au premier moment. Dans une quatrième partie nous exposons les résultats de simulation qui montrent d'une part les différences entre le champs moyen et l'IBM permettant d'évaluer l'impact de la structuration de la population, et d'autre part les différences entre le modèle approximé et l'IBM permettant d'évaluer la capacité du modèle approximé à prendre en compte cette structuration. Nous terminons par une discussion sur l'intérêt et les limites d'une telle méthode.

Simulations récursives

Aujourd'hui certains MSD particuliers utilisent eux-mêmes des simulations d'autres MSD dans le calcul de leur comportement. C'est le cas des modèles contenant des agents anticipatifs par exemple, où les agents simulent leur propre futur [GS05]. Un modèle d'agent anticipatif comprend une fonction qui simule un modèle de la représentation du monde portée par l'agent. Le fait d'appeler un moteur de simulation dans le cours d'une simulation est appelé la simulation récursive.

2.4 ρ_o : l'observation et le contrôle de systèmes complexes

L'observation des systèmes épidémiologiques, sociologiques et écologiques passe par l'utilisation de modèles statistiques qui permettent de synthétiser ou d'extrapoler des données d'observation. Nous allons décrire ces modèles dans cette section, notons qu'il ne s'agit pas de modèles du système cible lui-même mais de modèles des données récoltées. C'est à dire que ces modèles ne se substituent pas dans notre réflexion à des processus du système cible. Ils se substituent à des données récoltées localement sur le système cible, en fournissant des statistiques qui sont plus synthétiques et éventuellement plus globales.

2.4.1 Les modèles statistiques

Les modèles statistiques permettent de synthétiser une caractéristique des individus d'une même population sous la forme d'une variable aléatoire. Une variable aléatoire est une variable dont la valeur résulte d'une mesure sur un processus aléatoire. Elle peut être continue (à valeur dans \mathbb{R}) ou discrète, (à valeur dans \mathbb{N} par exemple). Si la variable aléatoire est une mesure effectuée sur un individu dans une population, le processus aléatoire est le choix aléatoire de l'individu dans la population. Le processus modélisé est bien le processus lié à l'observation qui consiste par exemple en épidémiologie à choisir l'animal à prélever dans un troupeau.

Un modèle statistique peut être vu comme une paire (Y, P) où Y est l'ensemble des observations possibles (les valeurs que peut prendre la variable aléatoire) et P la distribution de probabilité sur Y . Les modèles statistiques se construisent généralement en inférant les propriétés d'une population sur la base de propriétés observées pour un échantillon. On observe la distribution des valeurs que prend Y dans l'échantillon et on suppose que cette distribution sera la même pour l'ensemble de la population. C'est pour cette raison que nous considérons les modèles statistiques comme des descriptions synthétiques des mesures effectuées lors d'observations. On remplace un ensemble de mesures par une statistique sur ces mesures. Nous classons également l'ensemble des modèles de régression dans les modèles statistiques. Il s'agit des modèles statistiques qui ne décrivent plus une mesure, mais une relation entre des mesures observées. L'exemple le plus simple est la régression linéaire qui va décrire la relation entre deux variables aléatoires continues X et Y sous la forme d'une fonction linéaire f de \mathbb{R} dans \mathbb{R} . Il existe donc $(a, b) \in \mathbb{R}^2$ tels que $\forall x, f(x) = ax + b$. La fonction f modélise la relation F qui existe entre les deux variables aléatoires X et Y . On note cette relation $Y \sim F(X)$ et on parle dans ce cas de variable expliquée pour Y et de variable explicative pour X . De manière générale, les paramètres de la fonction de relation f sont conventionnellement notés θ . Pour la relation linéaire évoquée ci-dessus par exemple, $\theta = (a, b) \in \mathbb{R}^2$. On note $f(x, \theta)$ la valeur de la fonction f de paramètre θ en x . L'inférence statistique consiste alors à trouver le paramètre θ qui permet de décrire au mieux la relation observée entre les valeurs de X et de Y observées dans l'échantillon. On parle d'ajustement de la fonction f aux données. On distingue deux grandes classes de méthodes d'ajustements. Les méthodes fréquentistes, et les méthodes bayésiennes. Une caractéristique intéressante des modèles statistiques est qu'ils permettent une quantification rationnelle de la différence entre les prédictions du modèle et les observations. Il s'agit de la vraisemblance des observations au modèle. Un modèle statistique permet de décrire la distribution des valeurs possibles d'une variable aléatoire X sous la forme d'une loi statistique D et de ses paramètres Θ , telle que $X \sim D(\Theta)$. Cette relation donne la probabilité d'une réalisation x de X que l'on note $d(x, \Theta)$. La vraisemblance d'un ensemble d'observations $\{x_i\}_{0 < i < N}$ est calculée comme la probabilité d'avoir ces N réalisations de X sachant que $X \sim D(\Theta)$. On la note de la manière suivante :

$$V(\{x_i\}_{0 < i < N} | \Theta) = \prod_{i=1}^N d(x_i, \Theta)$$

Les sections 2.4.2 et 2.4.3 présentent des exemples de modèles statistiques utilisés en surveillance épidémiologique.

2.4.2 Les outils statistiques d'estimation de prévalence

En épidémiologie descriptive, l'outil de collecte de données est *l'enquête épidémiologique*. Nous considérons la composante dite *active* des réseaux de surveillance épidémiologique qui nous intéressent comme des systèmes d'enregistrement permanents de données issues d'enquêtes épidémiologiques descriptives. De nombreuses méthodes d'échantillonnage ont été définies pour être utilisées dans ces enquêtes [TDS⁺96, DMS10]. Des modèles statistiques simples donnent les tailles d'échantillons néces-

saires pour chacune des méthodes, ainsi que les intervalles de confiance sur les prévalences estimées.

En plus des méthodes classiques on peut trouver des travaux de recherche sur le développement de méthodes d'échantillonnages plus évoluées utilisant les caractéristiques propres aux phénomènes de santé recherchés. Le but étant de réaliser un échantillonnage ciblé sur différentes sous-populations dont on sait qu'elles ont des sensibilités différentes à la maladie (supérieures à la moyenne) et d'arriver à inférer la prévalence dans la population malgré le fait que l'échantillon ne soit pas représentatif de cette population [Can09, WEW09].

Dans tous les cas, ce sont des méthodes d'instantanés qui fournissent les informations recherchées à chaque campagne d'échantillonnage.

2.4.3 Approches statistiques pour le monitoring et la surveillance

Des méthodes de traitement d'observations beaucoup plus élaborées existent. Elles restent des approches statistiques mais utilisent des modèles statistiques si sophistiqués que la distinction entre les données modélisées et les processus du système cible qui génèrent ces données n'est plus très claire.

Ces recherches viennent du fait qu'un réseau de surveillance prélève une information indirecte, comme la sérologie ou le nombre de décès par exemple. Elles consistent à développer des algorithmes pour la détection d'aberrations dans les données issues des réseaux d'épidémiosurveillance. Le but est de travailler sur les séries temporelles issues des réseaux, de les modéliser avec des modèles statistiques calibrés en ligne grâce à des régressions, et d'identifier les données invraisemblables selon ces modèles pour les considérer comme des épidémies. On trouve plusieurs types d'algorithmes, plus ou moins accessibles aux néophytes en statistiques [HPH09]¹. Dans leurs versions les plus complètes, les algorithmes permettent de prendre en compte le fait que les données sont dépendantes, notamment au niveau spatial et temporel. Par exemple dans [PHT08] les auteurs considèrent que l'espérance du nombre de cas observés est la somme de deux termes, le premier correspond aux cas dûs au caractère endémique des symptômes (une sorte de bruit de fond plus ou moins stable) et le second aux cas qui s'ajoutent à ce caractère endémique et reflètent donc le départ d'une épidémie. L'équation 2.3 présente l'expression du terme épidémique. On peut y voir la prise en compte de l'aspect temporel par la relation entre l'état au temps t et l'état au temps $t-1$, et de l'aspect spatial par la relation entre l'état au temps t dans la zone « i » et l'état au temps $t-1$ dans l'ensemble des autres zones ($j \neq i$). L'ajustement d'un tel modèle aux données nous permet donc non seulement de décrire la cinétique de propagation du caractère observé, mais également une éventuelle structure spatiale observée au cours de la diffusion et qui se traduirait par les valeurs des coefficients ω_{ij} .

$$\zeta_{i,t} = \lambda_i \gamma_{i,t-1} + \phi_i \sum_{i \neq j} \omega_{ij} \gamma_{j,t-1} \quad (2.3)$$

Où,

- $\zeta_{i,t}$ est le nombre espéré de cas « épidémiques » dans la zone « i » au temps t ;

1. Notons qu'un package R qui les implémente a été développé [Höh07]

- $\gamma_{i,t}$ est le nombre espéré de cas dans la zone « i » au temps t ;
- λ_i et ϕ_i sont respectivement les coefficients déterminant l'origine locale ou externe des cas observés dans la zone « i » ;
- ω_{ij} est le coefficient représentant l'intensité de connexion entre la zone i et la zone j $j \neq i$.

2.5 ρ_m : L'utilisation des modèles

2.5.1 Résultats analytiques

La formalisation mathématique de certains systèmes nous permet d'utiliser les méthodes d'analyse mathématique pour étudier le comportement de ces systèmes. Nous avons dit dans notre présentation de la théorie systémique (section 2.1.1) que le point de vue systémique nous permettait d'uniformiser un discours sur les systèmes dynamiques en général. C'est à dire que certains concepts caractérisant un « type de comportement » (comme par exemple « être à l'équilibre ») peuvent être formalisés mathématiquement. Si le comportement dynamique du système est lui même formalisé de manière mathématique, la réponse à une question du type « est ce que ce système va avoir ce type de comportement ? » peut être obtenue de manière purement mathématique, par la résolution d'un système d'équations par exemple. Nous appelons ce type de démonstration un résultat analytique. Avant de poursuivre sur l'utilisation qui est faite de ces résultats analytiques, illustrons ces propos par un exemple en épidémiologie que nous allons dérouler tout au long de la thèse.

Nous considérons un système fermé (sans entrée ni sortie) dont le comportement dynamique est donné par le modèle SIS présenté section 2.3.2, formalisé sous la forme d'un système de deux équations différentielles (cf. équations 2.1, 2.2 page 30). Le « type de comportement » que nous cherchons à caractériser est la présence et la nature d'attracteurs. Un attracteur étant un équilibre stable, c'est à dire un point de l'espace d'états du système vers lequel le système va converger quel que soit son état initial à partir du moment où l'état initial du système est situé dans une certaine sous partie de l'espace d'états que l'on appelle alors « bassin d'attraction ». La question que l'on se pose est de savoir si pour un jeu de paramètres $(\beta, \gamma) \in \mathbb{R}^2$ il existe un attracteur, si oui lequel et quel est son bassin d'attraction ?

Comme on l'a dit plus haut, l'état du système SIS est constitué d'une paire de réels que l'on note $(s, i) \in \mathbb{R}^2$. Or on sait que $s = 1 - i$. Un équilibre stable est donc atteint si et seulement si la relation suivante est respectée :

$$\frac{di}{dt} = 0$$

Or, si l'on substitue s par $1 - i$ dans l'équation 2.2 et si l'on pose $R_0 = \beta/\gamma$, on obtient la relation suivante :

$$\frac{di}{dt} = (\beta - \beta i - \gamma)i = \beta i((1 - 1/R_0) - i) \quad (2.4)$$

L'état d'équilibre (noté (s^*, i^*)) est donc obtenu en mettant l'équation 2.4 à zéro. L'équation obtenue nous donne la solution unique suivante pour le point d'équilibre

(valable uniquement si $R_0 > 1$) :

$$(s^*, i^*) = (1/R_0, 1 - 1/R_0) \quad (2.5)$$

De plus, on observe que si $i(t) > i^*$ alors, on a
 $1 - 1/R_0 < i(t)$
 $\Rightarrow ((1 - 1/R_0) - i(t)) < 0$
 $\Rightarrow di/dt(t) < 0$ d'après l'équation 2.4.

Ce qui signifie, que la fonction $i(t)$ est strictement décroissante. Or on peut montrer que $\lim_{t \rightarrow \infty} di/dt(t) = 0 \Rightarrow \lim_{t \rightarrow \infty} i(t) = i^*$. On a donc démontré mathématiquement que l'état d'équilibre i^* va être rencontré. De la même manière, si $i < i^*$, on peut montrer que la fonction $i(t)$ est strictement croissante et donc que l'on va rencontrer l'état d'équilibre i^* .

Nous avons là un résultat analytique intéressant qui est le suivant : « si $R_0 > 1$, le système SIS admet un attracteur unique (s^*, i^*) (cf. équation 2.5) dont le bassin d'attraction est l'espace d'état du système, c'est à dire $\{(s, i) \in [0, 1]^2 | s + i = 1\}$ ». À titre d'illustration, on peut observer sur la figure 2.2 les trajectoires d'états d'un modèle SIS initialisé à trois positions différentes, et convergeant toutes vers l'attracteur (s^*, i^*) .

On remarque que la position de l'attracteur dépend uniquement du rapport $R_0 = \beta/\gamma$. On peut montrer que ce rapport a une influence importante sur le comportement du système. Nous en reparlerons dans notre chapitre d'application.

2.5.2 Comprendre pour contrôler

Prenons un exemple en épidémiologie où des MSD sont utilisés pour comprendre des points clefs de la propagation des épidémies en vue de les contrôler. Beaucoup de travaux de modélisation en épidémiologie s'intéressent à l'impact de la structure des populations sur la propagation des maladies. Ces travaux utilisent des modèles de graphes. Soit des graphes de contact comme [PCS⁺09] qui discute plusieurs méthodes de construction de ces graphes, soit des graphes de proximité comme [HXP⁺09] qui représentent par des graphes les risques de contamination via les marchés. L'article [MLPSV09] est une revue récente de ce qui est fait dans ce domaine en épidémiologie vétérinaire. Ces travaux ne sont pas spécifiquement dédiés à l'épidémiosurveillance. Cependant il est certain que l'analyse des réseaux sociaux fournit des indicateurs intéressants. Elle donne un certain nombre d'outils d'analyse de graphes représentant des réseaux sociaux. Elle est utilisée en épidémiologie, par exemple dans [ERE09] en santé humaine, où les auteurs calculent différents indicateurs pour déterminer les « nœuds à risque » et testent différentes stratégies de vaccination (sur des nœuds choisis selon leurs différents classements), sur différents types de scénarios (différentes valeurs de $\frac{\tau}{g}$ et différentes formes de réseaux) avec τ la probabilité de transmission et g le taux de guérison. Ils montrent que dans tous les cas ils obtiennent de meilleurs résultats qu'avec une vaccination aléatoire.

2.5.3 Les cadres expérimentaux

Dans le cas des MSD, il est possible d'associer un cadre expérimental à l'utilisation qui est faite d'un modèle. Ce cadre expérimental décrit les expériences (les

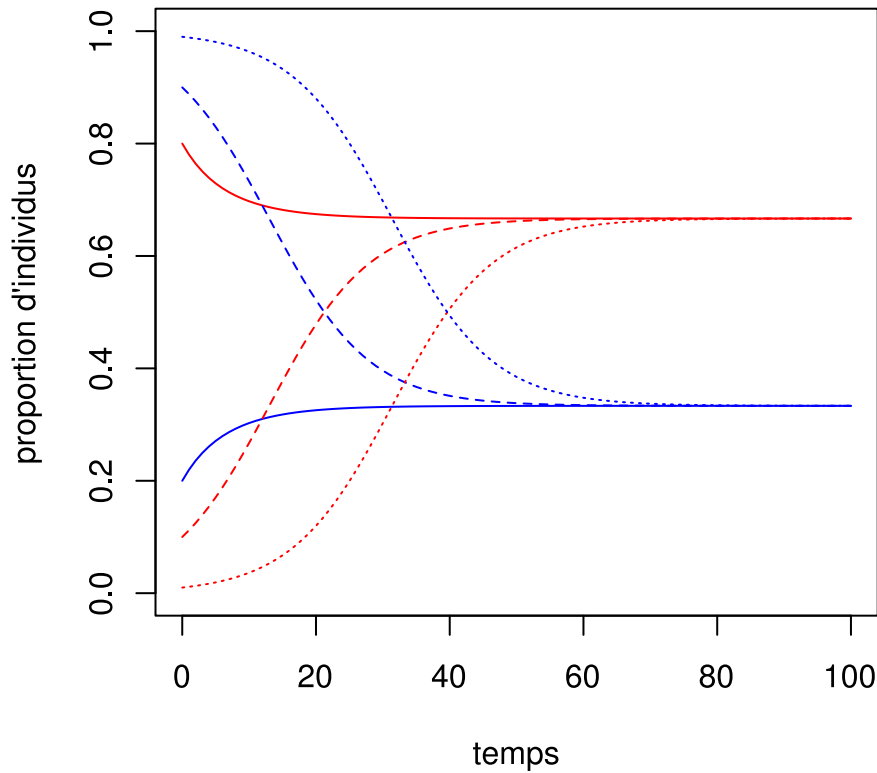


FIGURE 2.2: Illustration d'un attracteur du système SIS $(\beta, \gamma) = (0.199, 0.0667)$.

Les trois couples de courbes (pleines, points, tirets) représentent l'évolution de la proportion d'infectieux (courbes rouges) et de susceptibles (courbes bleues) pour trois états initiaux différents. On observe que dans les trois cas, le système se trouve en fin de simulation au même point d'équilibre ($s = 0.3$, $i = 0.7$). Il s'agit de l'attracteur (s^*, i^*) .

simulations) qui vont être réalisées sur le modèle dans le cadre de l'utilisation qu'on en fait. Le concept de cadre expérimental est utilisé dans le cadre de modélisation et de simulation de la triade de Minsky que nous proposons. Il nous vient de la TMS et il est en conséquence présentée en détail dans la section 2.7 qui présente la TMS comme le cadre général que nous utilisons pour développer nos outils.

2.5.4 Calibration

On parle aussi d'estimation de paramètres. L'estimation des paramètres d'un MSD consiste à identifier les valeurs de paramètres qui permettent au modèle de prédire au mieux les observations faites sur l'objet à modéliser. Nous distinguerons les méthodes globales et les méthodes itératives. Les premières estiment les paramètres à partir de l'ensemble des observations disponibles à une date donnée. Les secondes calculent une estimation sur la base des premières observations, puis la

Critère	Expression
Moindres carrés	$S(\theta, (y(t_i))_{i=1..N}) = \sum_{i=0}^{i=N} (y(t_i) - s(t_i, \theta))^2$
Écart moyen	$Mo(\theta, (y(t_i))_{i=1..N}) = \frac{1}{N} \sum_{i=0}^{i=N} \ y(t_i) - s(t_i, \theta)\ $
Écart maximal	$Ma(\theta, (y(t_i))_{i=1..N}) = \max_{i=0..N} (\ y(t_i) - s(t_i, \theta)\)$

TABLE 2.1: Mesures de distance pour les méthodes d'estimation globales. On note $y(t_i)$ l'observation mesurée à la date t_i sur le système source et $s(t_i, \theta)$ la valeur simulée à la date t_i en utilisant le jeu de paramètre θ .

corrigent de manière itérative à chaque nouvelle observation. On parle aussi d'assimilation de données pour ces deuxièmes qui assimilent petit à petit les observations dans le modèle. Nous n'utilisons dans cette thèse que des méthodes globales et nous ne présenterons donc que celles-ci.

Les méthodes globales peuvent être vues comme des problèmes d'optimisation. On va chercher les valeurs de paramètres qui permettent d'optimiser un certain critère de correspondance avec les mesures observées sur le système source (le système à modéliser). Imaginons que l'on dispose d'une série temporelle de N mesures de la sortie y , observées à des dates $\{(t_i)_{i=1..N} | i < j \Rightarrow t_i \leq t_j\}$ du système source. Notons $(y(t_i))_{i=1..N}$ cette série temporelle. L'objectif est de minimiser la distance entre ces observations et les observations générées par notre modèle. Différents calculs de distances sont présentés dans le tableau 2.1. Cette distance est le critère à optimiser. Pour un calcul de distance D (comme par exemple S , Mo ou Ma dont on donne l'expression dans le tableau 2.1) et un ensemble d'observations $(y(t_i))_{i=1..N}$, le problème d'optimisation consiste alors à trouver le jeu de paramètres θ qui minimise la fonction $D(\theta, (y(t_i))_{i=1..N})$.

2.5.5 Évaluation des systèmes de surveillance par simulation

On peut également trouver des travaux de modélisation et simulation intéressants sur l'évaluation de réseaux d'épidémiosurveillance de manière théorique. Citons l'exemple de [YTNNK08] qui teste des stratégies de surveillance de l'apparition d'une maladie au Japon. Les auteurs utilisent un modèle épidémiologique individu centré, définissent des indicateurs d'évaluation de la surveillance (temps avant détection, nombre d'individus déjà infectés avant la première détection) et estiment les valeurs de leurs indicateurs par simulation. Cet exemple nous intéresse tout particulièrement parce que ces indicateurs permettent une évaluation quantitative de différents systèmes de surveillance et cette évaluation se base sur les conséquences d'un choix de système de surveillance et non pas sur sa simple aptitude à détecter ou observer la maladie.

2.6 B : Processus de décision

2.6.1 Théorie de la décision

La théorie de la décision permet de formaliser (et de modéliser) la procédure de choix d'un décideur en fonction des préférences du décideur sur les conséquences

de ses choix. La théorie de la décision permet d'étendre la formalisation du choix dans l'incertain aux cas où le décideur a une vision subjective du risque d'occurrence des conséquences de ses choix et peut les classer de manière qualitative [Sav54]. Depuis, la théorie de la décision a également été utilisée pour formaliser la décision multi-critère ou multi-agents [AR86]. Notons que la théorie des jeux étudie depuis longtemps la prise de décision dans le cas multi-agent [vNM47].

2.7 Cadre général : la théorie de la modélisation et de la simulation

La Théorie de la Modélisation et de la Simulation (TMS) [ZKP00] fournit tout d'abord un cadre formel pour spécifier les modèles de systèmes dynamiques et leurs contextes. Elle fournit également une grille de lecture pour analyser les processus de modélisation des systèmes dynamiques en mettant en relation le modèle avec les objectifs de modélisation, le système représenté et les simulations. Cette théorie n'est pas le seul cadre existant pour décrire une activité de modélisation de systèmes dynamiques. On peut la comparer à d'autres efforts d'unification du vocabulaire et des concepts liés à la modélisation des systèmes complexes. La communauté de modélisation en écologie par exemple, qui est une communauté scientifique internationale et hétérogène, développe depuis plusieurs années un protocole de description des modèles individus centrés liés à l'écologie appelé Overview, Design concepts, Detail (ODD) [GBB⁺06, GBD⁺10]. La TMS a cependant les avantages que nous n'avons pas trouvés ailleurs d'être totalement générique, de se baser directement sur la systémique et de fournir un cadre formel et une sémantique opérationnelle pour la spécification des MSD et de leur simulation.

Les outils et les réflexions développés dans cette thèse sont illustrés par des exemples en épidémiologie mais sont destinés à une intégration générique dans la TMS. Aussi est-il indispensable que le lecteur se familiarise avec les concepts principaux de cette théorie.

2.7.1 Systèmes et niveaux de spécification

La section 2.1.1 résumait la pensée systémique, il est maintenant nécessaire de donner ici une définition plus précise d'un système. Étant entendu qu'un système représente une portion de la réalité, il nous est possible de décrire ce système à différents niveaux de détails suivant la place que prend ce système dans nos travaux de recherche. Zeigler parle de niveau de spécification pour désigner ces différents niveaux de détails. Il distingue sept niveaux de spécification de plus en plus précis. Au niveau 0 on décrit uniquement les domaines d'entrée/sortie du système, au niveau 1 on donne une liste de paires d'entrée/sortie observées sur le système, au niveau 2 on donne une fonction qui, pour un état initial donné donne la sortie unique qui sera observée sur le système pour chaque entrée possible, etc. L'annexe B reprend le chapitre 5 du livre présentant la TMS [ZKP00] en l'illustrant sur un exemple en épidémiologie.

2.7.2 Les entités de base de la TMS et leurs relations

Zeigler donne une définition précise des entités intervenant dans la TMS [ZKP00, Chap. 2] ainsi que des principales relations qui nous intéressent entre ces entités. Comme nous allons les utiliser tout au long de la thèse il est important d'en rappeler les définitions ici, ainsi que certaines subtilités apportées à ces entités depuis lors. Naturellement, de nombreux travaux de modélisation qui ne se réclament pas de la TMS s'intéressent aux mêmes entités. Nous en citerons quelques uns pour donner du relief à nos définitions mais l'objectif de ce paragraphe reste de fournir au lecteur des définitions claires des concepts que nous allons utiliser sans avoir la prétention de situer ces définitions dans un paysage exhaustif.

Le système source

Le système source est l'objet que l'on veut modéliser. Il s'agit d'un système, et il peut donc être défini à n'importe quel niveau de spécification. Il est généralement défini au niveau 1 sous la forme d'une relation d'entrée/sortie. Le système source tire son nom de ce cas général où il est considéré comme une boîte noire qui fournit une source de données de couples d'entrée/sortie que le modèle doit pouvoir reproduire.

Le modèle

Dans le cadre de la TMS, un modèle est défini de la manière suivante :

Un modèle de simulation [...] est un ensemble d'instructions, de règles, d'équations, ou de contraintes permettant de générer un comportement d'entrée/sortie.

On note que Zeigler parle de « modèle de simulation », nous utiliserons le terme de Modèle de Système Dynamique (MSD) pour désigner cet objet car dans certaines disciplines, un modèle de simulation peut être compris comme un modèle de système fermé² qui n'a pas de solution analytique. Entendons par là qu'on ne peut trouver de fonction analytique $f : T \rightarrow Q$ qui donne l'état du modèle en fonction du temps mais que l'on doit calculer la valeur de l'état du modèle de manière itérative à partir de sa valeur à une date t . Autrement dit, il est nécessaire de « simuler » le modèle à partir d'un temps initial t_0 pour lequel on connaît son état. On remarque qu'un système d'équations différentielles est bien un MSD. Cette définition de ce qu'est un « modèle » spécialise pour le cas dynamique la définition de Minsky donnée plus haut. L'objet dont on parle ici est bien le cas particulier du modèle de système dynamique ou autrement dit du modèle d'un processus.

Le simulateur

Défini comme ci-dessus, le modèle nécessite un agent capable de suivre les instructions données par le modèle de manière à générer son comportement. Il s'agit du simulateur. Nous utiliserons la définition de Zeigler qui est la suivante :

Un simulateur est un système de calcul (un micro-processeur, un réseau de micro-processeurs, l'esprit humain, ou de manière plus abstraite un

2. Un système fermé est un système sans entrée ni sortie, c'est à dire de la forme $SF = \langle T \times Q \times \Lambda \rangle$

algorithme) capable d'exécuter un modèle et de générer son comportement.

Le cadre expérimental

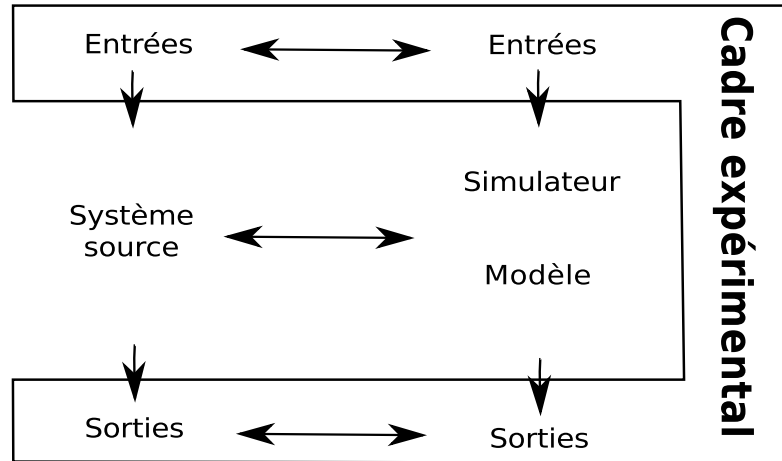


FIGURE 2.3: Cadre général de la modélisation et de la simulation selon Zeigler [ZKP00, Aum07]. Les flèches horizontales représentent les relations d'équivalence et d'approximation qui existent entre les entités du cadre. Les flèches verticales représentent un flux de données. La superposition du simulateur et du modèle illustre une relation d'équivalence particulière, un morphisme, que nous discutons dans le texte. (D'après [DBQsp].)

Pour relier le système source au modèle, la TMS associe la question posée à un Cadre Expérimental (CE), c'est-à-dire à un contexte dans lequel le comportement du système source nous intéresse. Ce cadre expérimental peut être décrit formellement à différents degrés de spécification (cf. tableau B.1) [Kli85, ZKP00]. Au niveau le moins précis, le cadre définit uniquement les domaines de définition des entrées et des sorties du système source (et donc du modèle associé). Au niveau le plus précis, il s'agit d'un ensemble de composants qui interagissent avec le modèle et qui nous permettent à la fois de simuler le contexte d'intérêt, et d'observer et analyser les trajectoires des variables d'états qui nous intéressent. De manière générale, le cadre expérimental doit décrire de manière rigoureuse l'ensemble des expériences qui peuvent être réalisées sur le modèle (ou sur le système source) dans le cadre de la question posée au modèle, ou de l'objectif de la modélisation. Une telle approche permet de relier rigoureusement la ou les questions posées au modèle (et au système source), à l'ensemble des expériences que l'on va réaliser (simulation du modèle ou expériences réelles sur le système source). On dit d'un cadre expérimental qu'il « peut être appliqué » à un modèle et réciproquement d'un modèle qu'il « s'accommode » à un cadre expérimental lorsque le modèle peut être forcé et observé selon les exigences du cadre expérimental.

Traore et Muzy, proposent une description formelle de cadres expérimentaux à différents niveaux de spécification [Tra06]. La formalisation de ces cadres nous permet d'appliquer un même cadre expérimental à plusieurs modèles et réciproquement

de tester un même modèle dans des cadres expérimentaux différents. Le premier cas nous permet par exemple de trouver le meilleur modèle ou un agencement préférentiel pour notre système si notre objectif est de concevoir un nouveau système. Le second cas nous permet de réutiliser un même modèle pour une question légèrement différente qui considérera le système étudié dans un contexte différent ou sous un point de vue différent. Dans leurs travaux, Traore et Muzy [Tra06] partent d'un objectif précis pour arriver à un cadre expérimental entièrement spécifié dans lequel ils peuvent tester des modèles différents pour choisir le mieux adapté. Dans leur exemple, ils cherchent un modèle de feu de forêt qui leur permette de prendre des décisions de lutte contre le feu en temps réel. Le cadre expérimental qu'ils définissent, simulé en même temps que le modèle, leur permet de tester différents scénarios d'occurrence des départs de feux tout en contrôlant les deux critères qui les intéressent : le temps d'exécution des simulations et l'erreur entre les résultats de simulation du modèle testé et ceux d'un modèle de référence. Le choix d'un modèle ou d'un autre ne va pas se faire au seul regard de sa faculté à imiter le système, mais sur un compromis entre les deux critères définis dans le cadre expérimental.

Une expérience consiste en pratique en un état initial et un ensemble de valeurs des paramètres du cadre expérimental. Par exemple, dans le cas de modèles de feux de forêts, la série temporelle des départs de feux est un paramètre du cadre expérimental. Il s'agit bien de la spécification d'une expérience qui aurait pu être réalisée sur le système. L'expérience consiste à déclencher des feux à certains moments à certains endroits. Le système correspond à la forêt. Les notions de cadre expérimental et d'expérience ainsi définies nous permettent de distinguer clairement les paramètres de l'expérience de ceux du modèle. Les discrétisations spatiales et temporelles choisies sont par exemple un paramètre du modèle.

Il est important de faire la différence entre plan d'expériences et cadre expérimental. Un plan d'expériences spécifie l'ensemble des valeurs prises par les paramètres et l'état initial des variables pour chaque simulation, ceci à la fois pour le modèle et pour le cadre expérimental. Il va consister en une liste de couples (cadre expérimental, Modèle), à tester. Un plan d'expérience est défini pour faire une analyse de sensibilité par exemple, ou optimiser un modèle au regard de certains critères définis a priori. Le plan d'expérience contient une spécification possible des trajectoires d'états d'entrée du modèle, le cadre expérimental devant quant à lui définir l'ensemble des trajectoires d'états admissibles en entrée dans le cadre d'une certaine question posée au modèle.

En TMS, la distinction claire, explicite, du modèle et de son contexte d'utilisation trouve son origine dans les sciences de l'ingénieur et leurs applications. Les objectifs de cette communauté sont de tester ou évaluer des protocoles ou des agencements de systèmes devant être conçus ou utilisés par l'industrie. Les ingénieurs peuvent donc décrire clairement les attentes qu'ils ont sur leurs systèmes, tout comme ils peuvent limiter leur domaine d'intérêt à des conditions d'utilisation restreintes. De plus, l'activité de modélisation et simulation en sciences de l'ingénierie (écriture du simulateur incluse) est vue comme la production d'un bien (le modèle et son simulateur) qui doit être réutilisable et qui doit donc être livré avec ses spécifications, c'est-à-dire avec son cadre expérimental général couvrant l'ensemble des cadres expérimentaux plus restreints (i.e. qui permettent moins d'expériences) pouvant être

dérivés selon la question posée par l'utilisateur du modèle³.

Notons enfin que le concept de cadre expérimental, une fois formalisé, permet un langage commun pour la spécification d'expériences virtuelles. La répétabilité, la distribution et l'automatisation des expériences scientifiques sont encore aujourd'hui des questions étudiées pour ce qui concerne les expériences réelles. Nous pouvons mentionner des travaux récents sur l'unification des modes de spécification des expériences [MRF⁺10] ou sur la distribution des protocoles expérimentaux [GBA⁺10]. Les simulations sont des expériences virtuelles réalisées sur des modèles de systèmes dynamiques. Ces expériences peuvent être spécifiés sous la forme de cadre expérimentaux. Cela permet notamment de construire des dépôts accessibles depuis internet qui contiennent des modèles et les cadres expérimentaux auxquels ils s'accommodent [CW09]. Notons de plus que le concept de cadre expérimental a aujourd'hui été introduit dans la communauté de modélisation en écologie [Aum07].

De la même manière que le modèle, le CE peut être spécifié à différents niveaux de spécification [Tra06].

2.7.3 La validation et la vérification

Le cadre expérimental permet de définir deux notions élémentaires de la modélisation et la simulation, la validation et la vérification, sous la forme de relations d'équivalences entre les entités de la TMS au sein d'un cadre expérimental. Ces relations d'équivalence (une entité peut remplacer l'autre dans le cadre expérimental) sont appelées des morphismes (cf. figure 2.3).

La validation

La validation concerne la relation de modélisation qui relie le système source au modèle. Elle consiste à substituer le système source au modèle à l'intérieur d'un cadre expérimental dans lequel le modèle est simulé et le système source est expérimenté. On compare alors les couples d'entrée/sorties que l'on peut générer pour les simulations du modèle et les expériences du système source. Un modèle dont le comportement serait indissociable du système source dans un cadre expérimental donné est un modèle valide du système source dans ce cadre expérimental. Le système source est considéré comme une boîte noire (plus ou moins transparente) et les informations concernant son fonctionnement (états et état initial compris) que l'on a pu mettre dans le modèle sont considérées comme des hypothèses de recherche ou des simplifications qui doivent être validées. On dit de deux systèmes équivalents à un niveau de spécification, qu'ils sont isomorphes à ce niveau de spécification. Zeigler distingue trois niveaux de validation :

- 1^{er} niveau : La validation de réplication consiste à montrer que deux systèmes sont isomorphes au niveau de spécification 1 (les mêmes couples d'entrée/sortie sont observés).
- 2^{eme} niveau : La validation de prédiction consiste à montrer que deux systèmes sont isomorphes au niveau de spécification 2, c'est à dire que le modèle doit reproduire des sorties que l'on n'a pas encore observées chez le système de

3. Ce paragraphe est tiré d'un article co-écrit pendant la thèse [DBQsp]. Cet article n'est pas encore publié mais il est disponible en Annexe de la thèse (cf. annexe G)

- manière à prouver que les hypothèses sur les états initiaux et les fonctions d'entrée/sortie sont suffisantes pour expliquer le comportement du système.
- 3^{eme} niveau : La validation structurelle consiste à montrer que deux systèmes sont isomorphes au niveau de spécification 3 ou supérieurs (on dit alors qu'ils sont homomorphes). Nous renvoyons le lecteur à l'ouvrage [ZKP00] pour une méthode opérationnelle permettant de montrer un tel homomorphisme.

La vérification

La vérification concerne la relation de simulation qui relie le modèle et le simulateur. Elle consiste à vérifier que le comportement simulé par le simulateur est bien celui que devrait avoir le modèle. La vérification se fait généralement au niveau de spécification 3 de manière à pouvoir vérifier si la succession des états du système respecte bien la fonction de transition d'états décrite dans le modèle et si les sorties générées par le simulateur respectent bien la fonction de sortie spécifiée dans le modèle. (On dit de deux systèmes équivalents à ce niveau de spécification qu'ils sont homomorphes.)

2.7.4 Une sémantique opérationnelle : DEVS

Dans la section 2.7.1, nous avons vu comment la TMS définit les systèmes dynamiques de manière générique à différents niveaux de détails. Par définition, un modèle de système dynamique est un ensemble de règles ou d'instructions qui doit nous permettre de générer un comportement d'entrée/sortie. Plusieurs formalismes, c'est-à-dire plusieurs cadres théoriques et syntaxiques, existent pour spécifier ces règles ou instructions. Ils sont associés à des méthodes ou algorithmes de simulation qui permettent d'utiliser les règles pour générer un comportement d'entrée/sortie. On dit qu'un formalisme A englobe un formalisme B, si pour tout modèle M_B exprimé avec le formalisme B, et pour tout cadre expérimental E_M applicable à M_B , il est possible de construire un modèle M_A avec le formalisme A qui soit homomorphe à M_B dans E_M . La TMS propose le formalisme Discrete Event System Specification (DEVS) qui englobe les deux formalismes les plus utilisés, à savoir DESS et DTSS que nous décrivons dans le paragraphe suivant. Dans la suite du document, nous dirons d'un formalisme englobé par DEVS, qu'il est conforme à DEVS (*DEVS compliant* en anglais).

Les formalismes de modélisation et les algorithmes de simulation

On peut distinguer deux grandes familles de formalismes pour représenter des comportements dynamiques, les équations différentielles qui représentent l'écoulement du temps de manière continue (on utilisera l'acronyme Differential Equations System Specification (DESS)) et la spécification de systèmes par événements discrets qui représente l'écoulement du temps de manière discrète (on utilisera l'acronyme DEVS). Notons le cas particulier de la deuxième famille où l'on considère l'écoulement du temps à intervalles réguliers (« pas de temps ») et que l'on appelle spécification de systèmes à temps discret (on utilisera l'acronyme Discrete Time System Specification (DTSS)). Historiquement, il n'était pas possible de coupler des modèles exprimés dans des formalismes appartenant à des familles différentes,

c'est-à-dire qu'un modèle devait nécessairement être exprimé entièrement soit en temps continu, soit en temps discret à pas régulier, soit sous la forme événements discrets. La TMS fournit les outils (et les démonstrations) permettant aujourd'hui de passer d'un formalisme à l'autre, et montre que le formalisme DEVS englobe les autres formalismes. Elle fournit également des algorithmes de « traduction » en DEVS d'un modèle formalisé dans un autre formalisme. Cela permet de définir des modèles dit multi-formalismes dont certaines parties sont spécifiées en temps continu et d'autres en temps discret mais qui, au final, sont toutes traduites en DEVS pour permettre leur simulation. Nous ne présentons ici que le formalisme DEVS car nous considérons les formalismes DESS et DTSS connus. Nous devons néanmoins introduire le formalisme de spécification de systèmes quantifiés (on utilisera l'acronyme Quantized System Specification (QSS)). Il consiste à discrétiser l'évolution de l'état du système (qui est donc une variable continue) selon un certain pas d'intégration, et à spécifier les intervalles de temps qui vont s'écouler entre deux pas quantiques de l'état. Ce formalisme et les intégrateurs associés permettent de spécifier de manière événementielle des systèmes qui sont initialement définis en temps continu [KJ01]. Nous utiliserons donc le formalisme QSS pour faire le lien entre les formalismes DEVS et DESS.

Le formalisme DEVS

Le formalisme DEVS provient d'un mode de spécification des systèmes dynamiques proposé par Zeigler. Ce mode de spécification est opérationnel, c'est à dire que la TMS propose un ensemble d'algorithmes qui permettent de simuler un système spécifié en DEVS. Pour rendre la spécification opérationnelle, on ne considère plus des segments d'entrée (qui sont rappelons le des fonctions d'une base de temps T vers un ensemble X de valeurs d'entrées possibles), mais des « événements discrets » qui sont composés d'une valeur de l'ensemble des entrées possibles, noté X . À une date $t \in \mathbb{R}$, un système DEVS peut ainsi recevoir un paquet d'événements qui est un sous-ensemble de X un peu particulier car un paquet peut contenir plusieurs occurrences d'un même élément de X . La fonction de transition d'état est en outre décomposée en trois fonctions de transitions d'états proprement dites δ_{ext} , δ_{int} , δ_{con} , et une fonction d'avancement du temps ta décrite ci-après. Nous renvoyons le lecteur à l'ouvrage de Zeigler [ZKP00], ou à la thèse de Raphaël Duboz [Dub04] pour une description précise des algorithmes de simulation et rappelons ici la formalisation d'un système DEVS pour le cas d'un système à un seul composant (on parle de modèle atomique) qui correspond au niveau 4 de la spécification des systèmes, et pour le cas d'un système multi-composants (modèle composé) qui correspond au niveau 7 de la spécification des systèmes.

Spécifier un comportement avec P-DEVS

Nous utilisons dans cette thèse une expression du formalisme DEVS dite « parallèle » qui se note P-DEVS pour « Parallel DEVS ». Ce mode d'expression permet, dans le cas d'un modèle composé, de ne pas spécifier une fonction d'ordonnancement des simulateurs à appeler mais de spécifier une possibilité de changements d'états simultanés (en parallèle) des sous systèmes. Dans ce cas, le comportement d'un modèle atomique est défini par la structure suivante :

$$DEVS = \langle X, Y, S, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda, ta \rangle$$

où

X est l'ensemble des événements d'entrée,

S est l'ensemble des états séquentiels (aussi appelé ensemble des phases ou d'états),

Y est l'ensemble des événements de sortie,

$\delta_{int} : S \rightarrow S$ est la fonction de transition interne,

$\delta_{ext} : Q \times X^b \rightarrow S$ est la fonction de transition externe, où X^b est un paquet d'événements de X ,

$\delta_{con} : S \times X^b \rightarrow S$ est la fonction de transition confluyente telle que $\delta_{con}(s, \emptyset) = \delta_{int}(s)$

$\lambda : S \rightarrow Y^b$ est la fonction de sortie,

$ta : S \rightarrow \mathbb{R}_0^+ \cup \infty$ est la fonction d'avancement du temps,

et où $Q = \{(s, e) | s \in S, 0 < e < ta(s)\}$, e est le temps écoulé depuis la dernière transition d'états.

Un modèle composé est spécifié de manière semblable au réseau de systèmes présenté dans l'annexe B (niveau 7 de la spécification de système) à la différence que les systèmes sont définis comme des structures P-DEVS. Le formalisme de spécification utilisé dans la thèse pour définir des modèles couplés est un peu différent car il s'agit d'une extension à DEVS. Il est présenté dans le paragraphe suivant.

Expliquons de quelle manière la structure P-DEVS décrit un comportement dynamique. Imaginons que le système arrive dans un état $s \in S$ à une date $t \in \mathbb{R}$, la fonction d'avancement du temps ta spécifie après quel intervalle de temps $ta(s)$ le système fera une transition interne. Trois cas de figure sont alors possibles. Si aucun événement externe n'est reçu avant $t + ta(s)$, la fonction de sortie spécifie le paquet de sortie $y = \lambda(s)$ observé en sortie du système à la date $t + ta(s)$. Si un paquet d'événements $x \in X^b$ est reçu à la date $t + e < t + ta(s)$, alors la fonction de transition externe spécifie l'état $s' = \delta_{ext}(s, e, x)$ dans lequel le paquet x entraîne le système. Si un paquet d'événements $x \in X^b$ est reçu à la date $t + ta(s)$ alors la fonction de sortie spécifie tout d'abord le paquet de sortie $y = \lambda(s)$ observé en sortie du système à la date $t + ta(s)$ puis la fonction de transition confluyente spécifie l'état $s' = \delta_{con}(s, x)$ dans lequel le système passe.

Simuler le comportement d'un modèle P-DEVS composé

Dans le cas d'un modèle composé, à une date t de la simulation, un coordinateur récupère les dates de prochains changements d'états prévues par les simulateurs des modèles composants, ainsi que toutes les sorties en provenance des simulateurs des modèles composants. Ensuite, il distribue les sorties et ordonne les transitions en fonction des dates prévues. La figure 2.4 présente une hiérarchie de modèles couplés, c'est à dire des modèles composés emboîtés les uns dans les autres (cf. annexe B). Les modèles sont couplés les uns aux autres via des ports d'entrée et de sortie (cf. annexe B.8 page 145). La figure 2.5 est une figure tirée du livre [ZKP00] qui illustre l'organisation de simulateurs pour ce modèle hiérarchique.

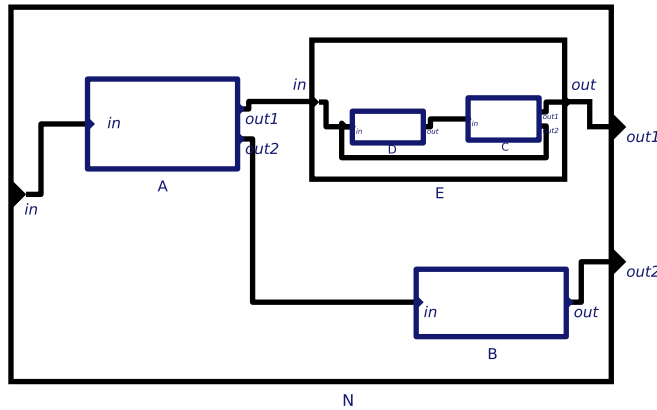


FIGURE 2.4: Hiérarchie de systèmes couplés.

N est un exemple de système couplé dont un des composants (le sous-système E) est lui même un système couplé.

Les extensions DS-DEVS et DSDE

Dynamic Structure Discrete Event System Specification (DS-DEVS) est une extension au formalisme DEVS apportée par F. J. Barros [Bar96] qui permet de spécifier un système couplé ayant une structure dynamique. C'est-à-dire tel que le nombre et la nature des modèles composants peuvent changer au cours du temps, ainsi que les connections entre ces composants. Nous renvoyons le lecteur à la page 235 du livre de Zeigler [ZKP00] pour une explication et une définition complète de la structure $DSDEVN_{\Delta}$ qui spécifie un réseau DEVS à structure dynamique. Le principe est d'associer à chaque modèle composé, un modèle particulier et unique appelé « exécutif » qui contient dans son état, le graphe des connexions internes et externes du modèle composé, ainsi que la liste de l'ensemble des modèles composants et de leurs définitions.

Notons que tous nos modèles DEVS sont spécifiés en Parallel-DEVS et nous utilisons en conséquence une version particulière de DS-DEVS adaptée à la spécification parallel-DEVS. Il s'agit de la spécification DSDE un peu plus récente et présentée dans l'article [Bar97]. Des exemples didactiques sont présentés dans [Bar03].

Propriété de fermeture sous couplage

Une propriété importante du formalisme DEVS est la propriété de fermeture sous couplage. Elle est démontrée dans [ZKP00] et assure que pour tout modèle DEVS couplé, il existe un modèle DEVS atomique qui lui soit isomorphe.

Implémentations existantes

Plusieurs plateforme de modélisation et de simulation DEVS sont disponibles. Citons PowerDevs [BK11] et Mimosa [Mül04]. Pour tous les développement évoqués dans cette thèse, nous utilisons la plateforme Virtual Laboratory Environment (VLE) [QDR09].

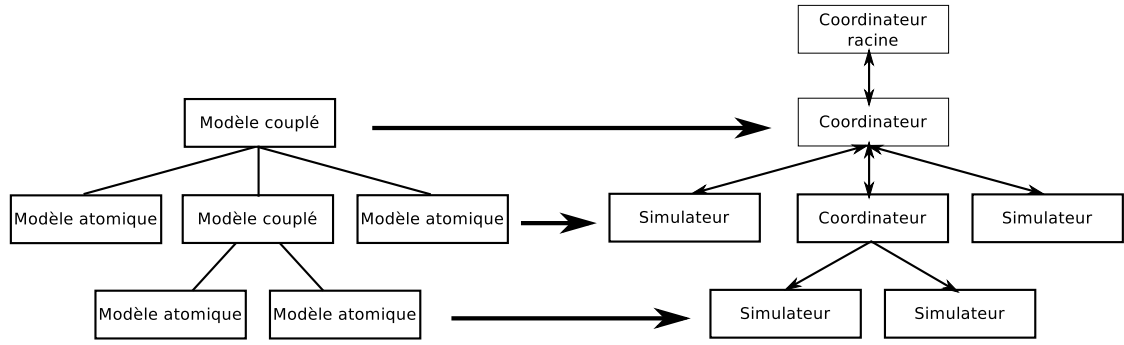


FIGURE 2.5: Mapping entre un modèle hiérarchique et un simulateur hiérarchique [ZKP00]. Un simulateur est construit pour chaque modèle atomique (cf. figure 2.4). Pour chaque modèle couplé, un coordinateur est défini qui a la même interface fonctionnelle qu'un simulateur. Un coordinateur parent va considérer de la même manière un simulateur et un coordinateur. Un coordinateur racine, situé tout en haut de la hiérarchie assure l'ordonnancement global des transitions d'états.

2.7.5 Conclusion : Retour sur la problématique

Fort des définitions exposées dans ce chapitre, nous pouvons préciser la problématique de la thèse. L'objectif de Minsky lorsqu'il proposait de considérer le triplet Modèle, Observateur, Objet, était de discuter de notre capacité d'introspection et de la place des modèles dans cette démarche. Notre objectif dans cette thèse est différent, il consiste à tester les méthodes explicites d'utilisation des modèles telle qu'elles sont décrites dans la TMS en application à des systèmes que l'on ne peut pas expérimenter.

En conséquence, cette thèse s'intéresse à un type de triade de Minsky particulier. Il s'agit des triades dans lesquels l'objet A peut être vu comme un système dynamique et le modèle A_B^* est un MSD. En nous plaçant dans le cadre de la TMS, nous ajoutons comme contrainte que A_B^* soit formalisé dans un formalisme conforme à DEVS. Cette contrainte n'est pas très restrictive puisque de nombreux formalismes sont conformes à DEVS. Nous considérons par ailleurs que la relation au modèle d'une triade T (ρ_m dans le chapitre précédent) consiste à réaliser des plans d'expériences sur A_B^* dans un cadre expérimental donné. Nous parlerons par la suite de « processus expérimental » pour désigner l'activité qui consiste à réaliser des plans d'expériences et qui peut être un processus dynamique en interaction avec d'autres processus comme celui d'observer le système. La relation à l'objet, en revanche (ρ_o dans le chapitre précédent), n'est pas à proprement parler une relation d'expérimentation dans le sens où l'observateur de Minsky ne peut pas réellement faire d'expériences sur l'objet. Il peut uniquement avoir des actions d'observation et de contrôle sur l'objet.

Chapitre 3

Proposition : Le modèle de la triade de Minsky et sa simulation

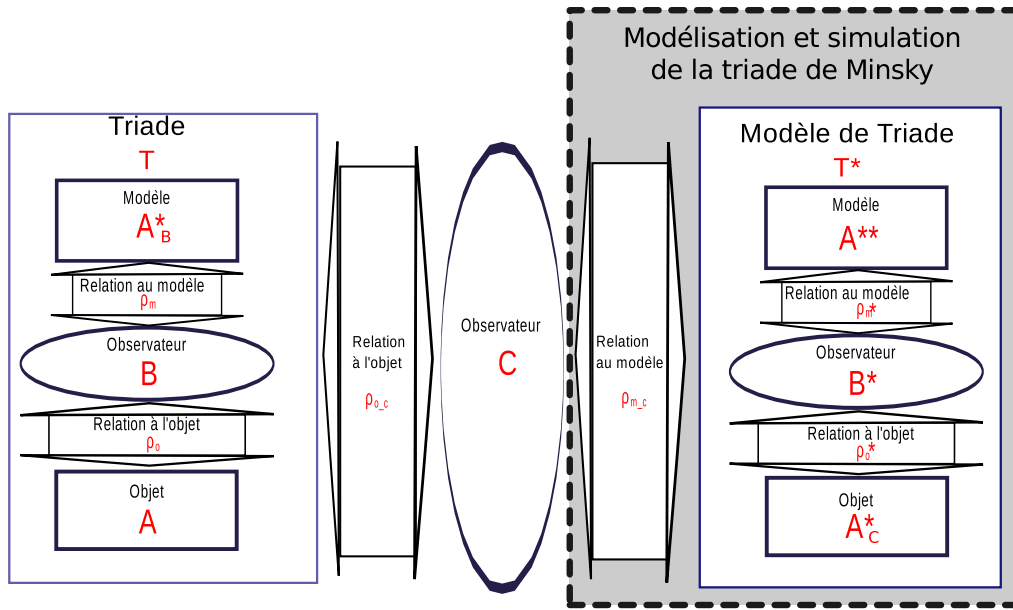
RÉSUMÉ

Nous proposons un cadre générique qui permet de construire un modèle simulable T^* d'une triade de Minsky T vue comme un système dynamique. Le premier composant du système Triade est le système cible. Nous explicitons quelques conditions sur la formalisation du modèle du système cible A_C^* . Le second composant est la relation au modèle. Nous verrons comment construire un modèle ρ_m^* de cette relation sous la forme d'un Modèle de Processus Expérimental (MPE) générique formalisé en DEVS. Un cas d'utilisation appliqué à la calibration de modèle a été développé. Nous verrons ensuite comment le MPE permet d'interagir avec le modèle A^{**} . Nous proposons également un modèle ρ_o^* de la relation de l'observateur à l'objet qui est vue comme un Système Pour l'Observation et le Contrôle (SPOC). Nous verrons comment ce composant peut être formalisé de manière opérationnelle sous la forme d'un ensemble de processus d'observation et de contrôle. Nous proposons des modèles génériques pour les processus d'observation et un schéma d'interaction entre les modèles composants du modèle de SPOC. Nous positionnerons enfin le modèle processus de décision B^* dans l'architecture obtenue. Nous obtenons un cadre modulaire pour spécifier le modèle T^* sous la forme d'un modèle DEVS couplé. Une proposition d'implémentation du modèle T^* dans la plateforme Virtual Laboratory Environment (VLE) permet son expérimentation, c'est à dire la relation ρ_{m_c} que l'utilisateur de notre cadre entretient avec le modèle T^* .

3.1 Introduction : le modèle du système Triade

La triade T présentée dans l'exposé de notre problématique (chapitre 1) constitue le cadre que nous proposons pour étudier l'interdépendance entre l'objet, l'observateur et le modèle de l'objet dans la triade de Minsky. Ce cadre consiste à construire

un modèle T^* de la triade T . Il est illustré figure 1.7. Le chapitre d'état de l'art (chapitre 2) a précisé le type de triade T qui nous intéresse. Il s'agit de triades dans lesquelles l'objet modélisé est un Socio-Éco-Système (SES). Le chapitre 2 a également expliqué pourquoi la théorie de la modélisation et de la simulation a été choisie pour modéliser la triade T en la considérant comme un système dynamique. Dans ce chapitre nous proposons une spécification modulaire d'un modèle de système triade en utilisant la théorie de la modélisation et de la simulation. L'objectif est de permettre la construction du modèle T^* et son expérimentation. Pour pouvoir expérimenter T^* , il est nécessaire d'implémenter un simulateur qui va nous permettre de le simuler. L'activité qui consiste à expérimenter le modèle T^* correspond à la relation ρ_{m_c} de la triade T' . La figure 3.1 identifie les entités de la triade T' considérées dans ce chapitre.


 FIGURE 3.1: Modélisation et simulation de la triade T^*

3.1.1 T : Le système triade

La triade de Minsky vue comme un système dynamique peut être vue comme un système composé de deux sous-systèmes : le modèle de système cible et le modèle d'observateur de Minsky vu comme un système cybernétique. En effet, en modélisant la triade de Minsky, nous proposons de modéliser le système composé du système cognitif S (qui contient l'observateur B et le modèle A_B^*) et de son environnement F (qui contient l'objet A). Nous appelons ce système : le système triade. Ce système correspondrait à l'univers $U = S + F$ évoqué dans la section 2.1.2 de l'état de l'art (page 25). Nous expliquons dans la section section 2.1.2 qu'un système cognitif S ne pouvait acquérir de connaissance objective sur l'univers U en raison de son appartenance à U . Notons qu'en modélisant l'univers $U = S + F$ nous nous plaçons dans une position d'observateur extérieur vis à vis du modèle de la triade. Nous pouvons ainsi expérimenter ce modèle tout en restant observateur omniscient et passif (l'observateur C présenté dans la figure 3.1).

De plus, nous proposons de décomposer la chaîne « pragmatique » du système cognitif (cf. section 2.1.2) en faisant apparaître explicitement l'opérateur d'observation, l'utilisation d'un modèle, et l'opérateur de décision. Le chapitre d'état de l'art a montré que cette décomposition est possible pour le type de triades qui nous intéresse. Nous nous plaçons dans le cas où nous nous questionnons justement sur des protocoles d'observation, des modes de modélisation et des modes de contrôle de l'objet de la triade de Minsky. Dans le contexte thématique auquel la thèse est appliqué, l'opérateur d'observation est un système de surveillance épidémiologique, les modèles sont des modèles épidémiologiques et l'opérateur de décision peut être spécifié sous forme d'une réglementation concernant les mesures prophylactiques à appliquer lors de la détection d'un cas de maladie par exemple.

Rappelons que nous nous intéressons à un cas particulier de triade de Minsky dans lequel les modèles utilisés (l'entité A_B^* de la figure 3.1) sont des modèles de simulation, c'est à dire des modèles que l'on peut expérimenter. Nous pouvons maintenant définir le système triade sous la forme d'un ensemble de sous-systèmes en interaction. Les différentes entités de la triade de Minsky y apparaissent mais ne constituent pas toutes des sous-systèmes du système triade. L'entité A constitue un sous système du système triade que nous appelons le système cible. L'entité B est décomposée en trois sous systèmes : le Système Pour l'Observation et le Contrôle (SPOC), le système de Processus Expérimental (PE) et le système de processus de décision. Le SPOC correspond à la relation ρ_o que l'observateur entretient avec l'objet (cf. figure 3.1). Il récolte des informations sur le système cible et agit éventuellement sur celui-ci par des actions de contrôle. Le système de processus expérimental correspond à la relation ρ_m d'utilisation du modèle par l'observateur (cf. figure 3.1). Il utilise des informations sur le système cible et le modèle A_B^* pour générer des informations d'intérêt sur le système cible à partir de résultats d'expériences sur le modèle. Enfin, le système de décision correspond au processus de décision qui consiste pour l'observateur B à modifier son SPOC en fonction des informations fournies par le processus d'expérimentation du modèle A_B^* . La figure 3.2 présente le système triade et la place que les entités de la triade de Minsky occupent dans le système triade.

Notons ici un point important : le modèle A_B^* n'est pas considéré comme un sous système du système triade parce qu'il ne peut pas être associé à un processus qui aurait lieu dans le système triade. Le comportement d'entrée/sortie décrit par A_B^* ne peut pas être mis en interaction directe avec le comportement d'entrée/sortie des sous systèmes du système triade. On peut se représenter le modèle A_B^* comme un ensemble d'équations se trouvant « dans la tête » de l'observateur de Minsky.

3.1.2 T* : Le modèle du système triade

Pour modéliser ce système nous allons considérer une hiérarchie de systèmes¹ légèrement différente qui nous semble plus adaptée. Comme nous l'avons dit dans notre présentation de la systémique section 2.1.1, les limites d'un système sont déterminées de manière subjective. La nouvelle configuration que nous proposons pour le système triade ne traduit donc pas un changement de nature du système triade mais plutôt

1. La notion de système hiérarchique est illustrée par la figure B.5 présentée en annexe (annexe B)

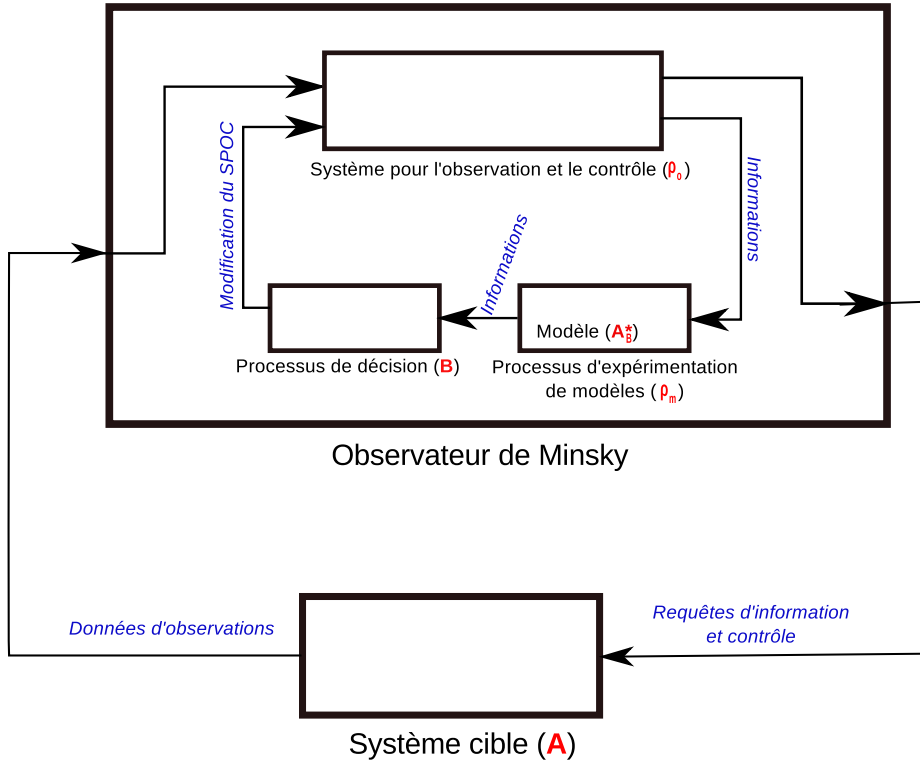


FIGURE 3.2: Le système Triade. Le système triade est composé d'un système cybernétique cognitif (l'observateur de Minsky) en interaction avec un système cible qu'il observe et qu'il contrôle éventuellement (l'objet A de Minsky). L'observateur de Minsky peut être décomposé en trois sous système : le SPOC, le processus de décision et le processus d'expérimentation de modèles, que l'on associe respectivement à la relation ρ_o que l'observateur de Minsky entretient avec A , à l'observateur de Minsky B , et à la relation ρ_m que l'observateur de Minsky entretient avec le modèle. Le modèle A_B^* est compris dans l'état du système PE. Les connexions entre les systèmes désignent le type de variables portées par les événements échangés entre les sous systèmes.

une manière différente de le voir. Cette réorganisation comprend deux modifications. Premièrement, le système cible et le SPOC sont décomposés en sous systèmes qui se trouvent à un même niveau hiérarchique que l'on nomme le système cible contrôlé. Cette première réorganisation est due au fait que les SES sont généralement composés d'entités que l'on peut distinguer et qui sont sujettes à des modes d'observation et de contrôle éventuellement différents (cf. sections 2.3.3 et 2.4). Deuxièmement, le processus d'expérimentation de modèles est sorti du système d'observateur de Minsky qui se retrouve réduit à un processus de décision. Cette dernière modification a pour unique objectif de favoriser la mise en correspondance entre le système triade et la triade de Minsky.

Rappelons que le modèle A_B^* est un Modèle de Système Dynamique (MSD). Il est donc modélisé lui même par un MSD : le modèle A^{**} qui se substitue au modèle A_B^* dans la réflexion de l'observateur C et peut donc être noté $(A_B^*)_C^*$. Notons qu'il peut s'agir du même objet que A_B^* ($A^{**} = A_B^*$) mais pas nécessairement (cf. section 1.4 page 18). Nous avons souligné plus haut que le système décrit par le

modèle A_B^* ne pouvait pas interagir directement avec les sous-systèmes de la triade. Pour la même raison, nous ne pouvons pas coupler le MSD A^{**} avec les modèles des sous-systèmes de la triade. Le modèle de processus expérimental (qui correspond à ρ_m^* dans la figure 3.1) modélise la relation ρ_m que l'observateur B entretient avec le modèle A_B^* . Rappelons la définition d'un MSD que nous avons donnée dans l'état de l'art (section 2.7.2 page 41) et qui est la suivante : « Un modèle de simulation [...] est un ensemble d'instructions, de règles, d'équations, ou de contraintes permettant de générer un comportement d'entrées/sorties ». On peut donc dire qu'un MSD décrit le comportement d'un système dynamique. Le MPE ne peut pas interagir directement avec le système dynamique décrit par le modèle A^{**} en lui envoyant des événements car le modèle A^{**} est un modèle de modèle et il reste donc un modèle (un ensemble d'instructions, de règles, d'équations,... et pas un système dynamique) au sein du modèle T^* . Pour réaliser des expériences qui nécessitent des interactions dynamiques avec le système dynamique décrit par le modèle A^{**} , le MPE doit construire un Cadre Expérimental (CE) (cf. section 2.5.3 page 37) qui est lui-même un MSD que le MPE va devoir simuler en même temps que le modèle A^{**} et qui va réaliser les interactions dynamiques entretenues avec le système dynamique représenté par le modèle A^{**} . Nous expliquons dans les sections 3.3 et 3.4 comment le MPE permet d'implémenter cette interaction particulière entre le système dynamique décrit par le modèle A^{**} et les MSD des sous-systèmes de la triade de Minsky.

La figure 3.3 présente la décomposition hiérarchique du modèle T^* .

3.1.3 Plan du chapitre

Nous présentons successivement dans ce chapitre : le modèle de système cible (A_C^*) ; le modèle de processus d'expérimentation de modèles (ρ_m^*) ; le modèle A^{**} et son cadre expérimental ; le modèle de SPOC (ρ_o^*) ; le modèle de processus de décision (B^*) ; et enfin les outils informatiques développés pour permettre l'expérimentation du modèle T^* (ρ_{m_c}). Pour chaque section, nous rappellerons la place de l'entité de la triade de Minsky considérée dans le modèle T^* en reprenant la figure 3.3.

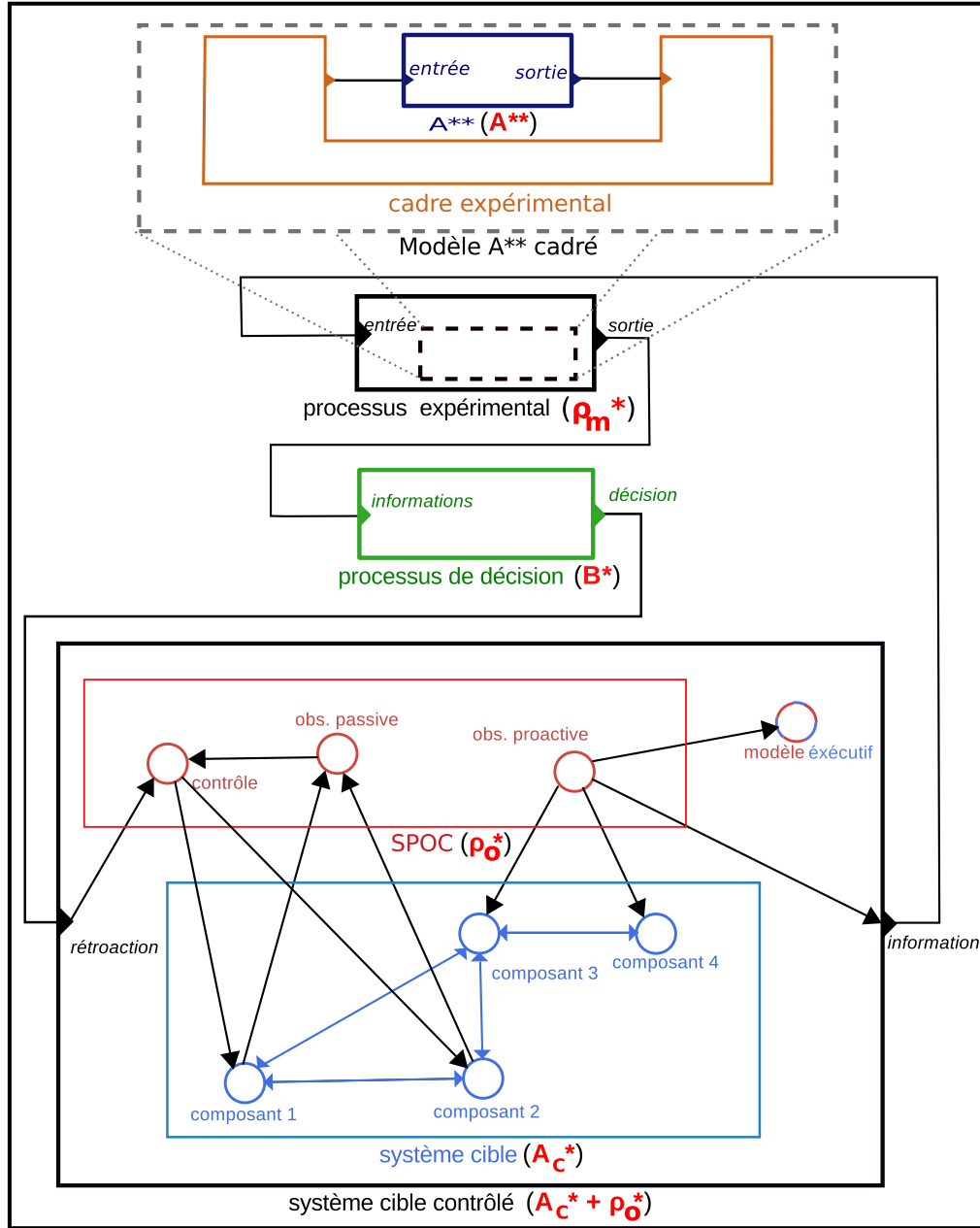
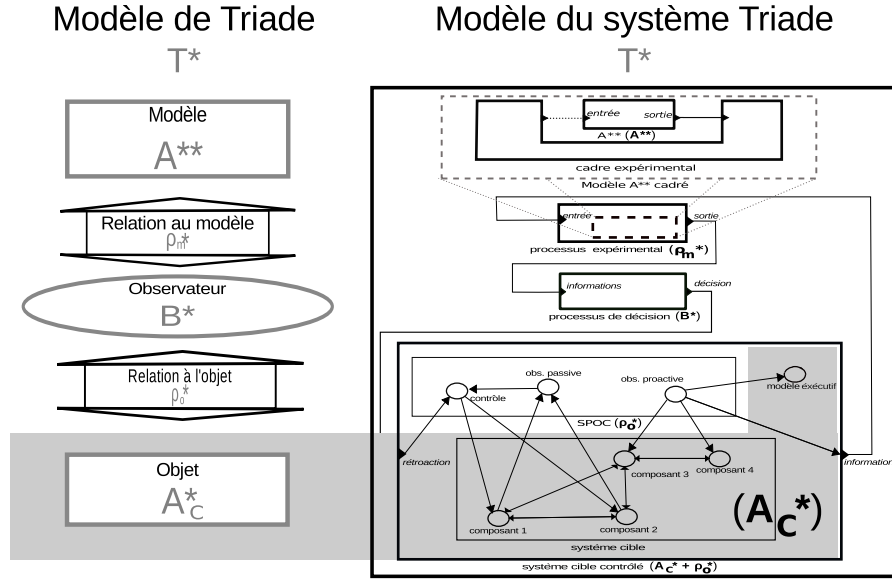


FIGURE 3.3: Structure du modèle T^* . Les boîtes rectangulaires et circulaires représentent des modèles de systèmes dynamiques, spécifiés comme des modèles atomiques ou des modèles composés. Les connexions sans flèches représentent des connexions « port \rightarrow port ». Les connexions avec des flèches représentent une connexion « modèle \rightarrow modèle » (c'est à dire une ou plusieurs connexions « port \rightarrow port »). Les modèles représentant des processus du système cible (exemple de structure en bleu clair) et les modèles représentant des processus du SPOC (exemple de structure en rouge), se situent au niveau hiérarchique du système cible contrôlé. Le modèle de processus de décision (en vert) se situe au niveau hiérarchique de la triade. Le modèle exécutf du modèle de système cible contrôlé (en rouge et bleu clair) assure la gestion des connexions dynamiques entre les modèles du SPOC et les modèles composants du modèle de système cible. Le modèle de processus expérimental (en noir) se situe au niveau hiérarchique de la triade. Il encapsule le modèle A^{**} et l'expérimente en utilisant un cadre expérimental (cf. sections 3.3 et 3.4).

3.2 A_C^* : Le modèle de système cible

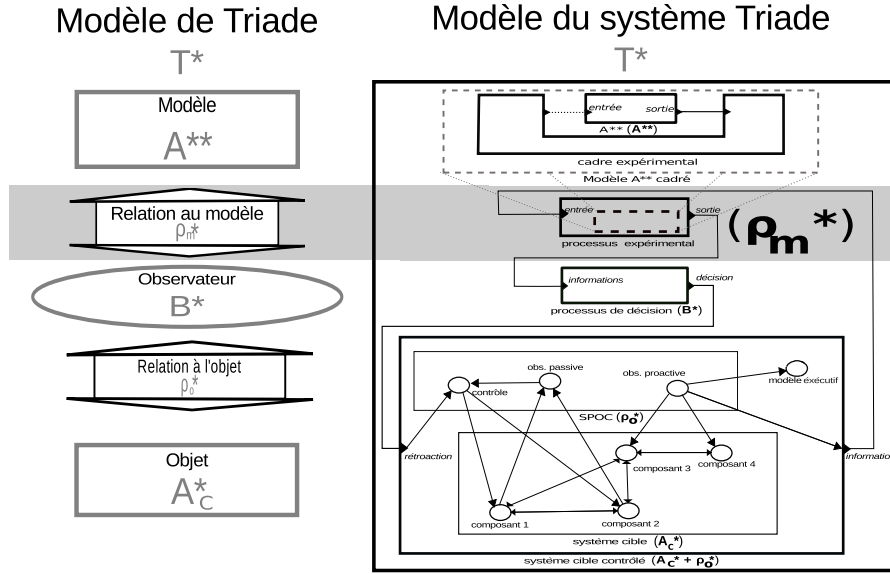


Il s'agit d'un modèle du système que l'on veut observer, modéliser, et éventuellement contrôler. Un des intérêts de disposer d'un modèle de la triade est que cela offre la possibilité de modifier le système cible et de tester la robustesse du système Triade à cette modification. Voyons les conditions que doit remplir un modèle de système cible pour pouvoir être intégré dans le modèle T^* .

Le modèle de système cible doit être spécifié en DEVS ou dans un formalisme conforme à DEVS (cf. chapitre 2). Il doit être spécifié sous la forme d'un modèle DEVS couplé à structure dynamique selon la spécification DS-DEVS présentée dans le chapitre 2. Cela nous permet de tester des hypothèses sur certains composants du système cible toutes choses égales par ailleurs. Cela nous permet également de jouer sur la structure du modèle de système cible sans en changer les composants (quels composants sont liés à quels autres, cf. chapitre 2). Enfin, il est possible de définir des classes de modèles composants, si nous pensons avoir à faire à des groupes de composants ayant les mêmes caractéristiques fonctionnelles. Il est ensuite possible de jouer sur la définition de ces classes. Nous présentons en annexe les outils offerts par le VLE pour spécifier ces classes de modèles (cf. Annexe A.1), et pour générer automatiquement un réseau de modèles à partir de la liste des classes de modèles composants et d'un graphe représentant la structure du réseau (Cf. Annexe A.2).

Le modèle de système cible doit également respecter un certain nombre de contraintes qui lui permettent une interaction avec le modèle de SPOC. Ces contraintes s'expriment au niveau de la spécification du modèle de système cible contrôlé (cf. figure 3.3), sous la forme d'un protocole d'interaction entre les composants du système cible et les composants du SPOC. Ces contraintes sont présentées en même temps que le modèle de SPOC car elles dépendent du type de composants utilisés dans le modèle de SPOC.

3.3 ρ_m^* : Le modèle de processus expérimental



3.3.1 Un modèle d'expérimentation de modèles

Toute notre proposition est en partie justifiée par le fait que les cadres expérimentaux (présentés dans l'état de l'art section 2.7.2 page 42), tout comme les modèles, peuvent être spécifiés avec le formalisme DEVS. Il est donc possible d'utiliser ce formalisme pour spécifier les expériences virtuelles à réaliser sur le modèle A^{**} . Le Modèle de Processus Expérimental (MPE) va spécifier et manipuler les expériences sur A^{**} . Cette technique est basée sur le principe de la simulation récursive telle que nous l'avons décrite dans l'état de l'art (cf. section 2.3.4). Nous utiliserons le terme de « simulation encapsulée » pour faire référence à cette technique de manière à souligner que le modèle A^{**} est différent du modèle de processus expérimental et donc que la technique de simulation encapsulée ne constitue pas un appel récursif à une même opération. Décrivons cette technique plus en détail. Le modèle A^{**} est lui-même simulé, ce qui implique deux niveaux de simulations. On appelle la simulation du modèle de processus expérimental (et donc des modèles de sous-systèmes de la triade) la « simulation de base » et la simulation de A^{**} la « simulation encapsulée ». Le MPE lance des simulations du modèle A^{**} et récupère les résultats. La figure 3.4 illustre la technique de simulation encapsulée. Les principales étapes sont les suivantes :

- Lorsque nécessaire, la simulation de base est stoppée,
- les spécifications d'expériences sont générées par la simulation de base,
- les simulations encapsulées sont lancées,
- les sorties de simulations encapsulées sont analysées,
- des résultats sont retournés à la simulation de base,
- la simulation de base continue.

On désignera les modèles simulés dans les simulations encapsulées sous le terme de modèles encapsulés. Le MPE doit être spécifié et implémenté de telle manière qu'il puisse être lui-même inclus sous la forme d'un composant modulaire dans n'importe quelle simulation.

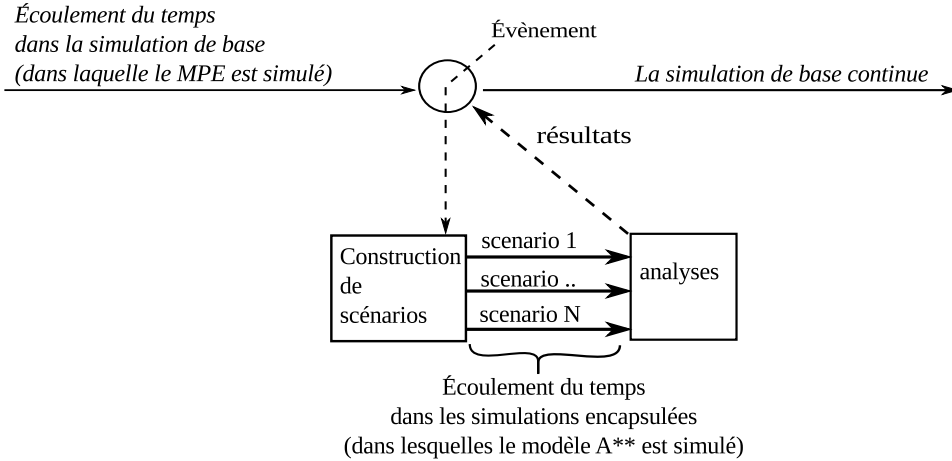


FIGURE 3.4: Principe de simulation encapsulée (d'après [GS05]). Les simulations encapsulées sont lancées depuis la simulation de base. Les résultats de simulations encapsulées sont utilisés par la simulation de base (voir section 3.3.1).

Le MPE lui-même est un MSD et le formalisme DEVS [ZKP00] est pertinent pour deux raisons. Tout d'abord, DEVS permet une approche hiérarchique et modulaire qui permet la décomposition du MPE en sous modèles composant et la ré-utilisation de certains sous-modèles composants pour des applications différentes. Ensuite, une fois le MPE spécifié en DEVS, il est possible d'utiliser toutes les propriétés d'intégration fournies par DEVS pour réaliser le couplage entre le MPE et les autres modèles de sous-systèmes de la triade.

3.3.2 Définition d'un protocole expérimental itératif

Avant de présenter le modèle de processus expérimental, présentons ce qu'il modélise, à savoir un protocole expérimental réalisé sur un modèle de système dynamique. Nous distinguons deux niveaux dans notre description d'un protocole expérimental. Le niveau le plus bas est « l'expérience », il spécifie notamment le modèle A_B^* expérimenté et son cadre expérimental. Le plus haut est le « protocole expérimental » et il correspond à la relation ρ_m que l'observateur de Minsky entretient avec le modèle. Le protocole expérimental décrit la manière de générer un ensemble d'expériences (que l'on appelle un plan d'expérience) et la manière d'analyser les résultats de ces expériences.

Nous définissons une expérience comme un cadre expérimental couplé à un modèle. Comme nous l'avons vu dans le chapitre 2, la notion de cadre expérimental est particulièrement importante pour étudier les relations entre systèmes. Dans notre cas nous n'utilisons pas les cadres expérimentaux pour comparer deux systèmes mais pour spécifier les conditions sous lesquelles le modèle est expérimenté. En conséquence, nous restreignons notre utilisation des cadres expérimentaux à la description suivante donnée par Zeigler :

« Un cadre expérimental consiste typiquement en trois modèles : Un modèle pour générer les entrées du modèle que l'on étudie, un modèle pour collecter les données de sorties du modèle, et un modèle pour contrôler

l'exécution de la simulation. Nous appelons ces trois modèles respectivement un générateur, un accepteur et un transducteur. Ces trois modèles sont analogues, dans leurs fonctions, à un générateur de signal, un oscilloscope et un analyseur de spectre, respectivement, dans une expérience sur un circuit électrique. » [ZKP00]

Nous sommes cependant conscients qu'un cadre expérimental peut prendre un aspect différent et que cette notion a une position importante dans la théorie et les activités de la Modélisation et la Simulation.

Dans cette proposition, nous considérons seulement des cadres expérimentaux définis au plus haut niveau de spécification, c'est à dire au niveau d'un modèle de système dynamique composé, et nous utilisons des CE qui spécifient une unique expérience. Nous n'avons en effet besoin dans cette première étude que d'une manière opérationnelle de spécifier le contexte dans lequel le modèle est expérimenté. Un cadre expérimental doit en fait comprendre les informations suivantes :

- un modèle de système dynamique appelé « Générateur » qui doit pouvoir générer les événements d'entrée à envoyer au modèle au cours de la simulation,
- un MSD appelé « Accepteur » qui détermine quelle variable du modèle expérimenté doit être observée et quand,
- éventuellement un MSD appelé « Transducteur » qui spécifie un ensemble de pré-traitements des données récoltées de manière à extraire des valeurs d'intérêt particulières,
- une condition d'arrêt pour la simulation (usuellement la durée de simulation mais il peut s'agir d'un autre critère).

Comme nous l'avons dit précédemment, nous définissons une expérience comme la combinaison d'un cadre expérimental et du MSD expérimenté. Le modèle expérimenté lui-même doit donc être spécifié. Quel que soit la nature du MSD utilisé, les éléments suivants peuvent être distingués : *i. les paramètres*, *ii. l'état initial*, et le reste de la spécification du modèle que nous désignerons comme étant *iii. la dynamique du modèle*. Dans le cadre de la modélisation de la triade de Minsky, le modèle expérimenté est le modèle A^{**} . Il est présenté ainsi que son couplage avec son cadre expérimental dans la section 3.4.

Nous pouvons maintenant définir un protocole expérimental comme étant composé de :

- la description d'un plan d'expérience,
- la manière d'analyser les résultats d'expérience,
- une procédure pour générer un nouveau plan d'expérience.

Le protocole expérimental peut ainsi être itératif si les résultats d'une expérience peuvent être utilisés pour définir l'expérience suivante. C'est le cas pour les protocoles d'optimisation par simulation où les valeurs de paramètres utilisées pour une simulation sont calculées en fonction des résultats d'expériences précédentes. Dans ce cas, il est nécessaire de disposer d'un *feedback* entre les résultats d'un plan d'expérience et le processus de génération du nouveau plan d'expérience. La figure 3.5 présente la composition d'un tel protocole expérimental itératif.

Ayant présenté comment le modèle A^* est expérimenté selon un protocole expérimental, nous pouvons maintenant définir le modèle de processus expérimental qui nous sert à modéliser la réalisation dynamique d'un protocole expérimental itératif.

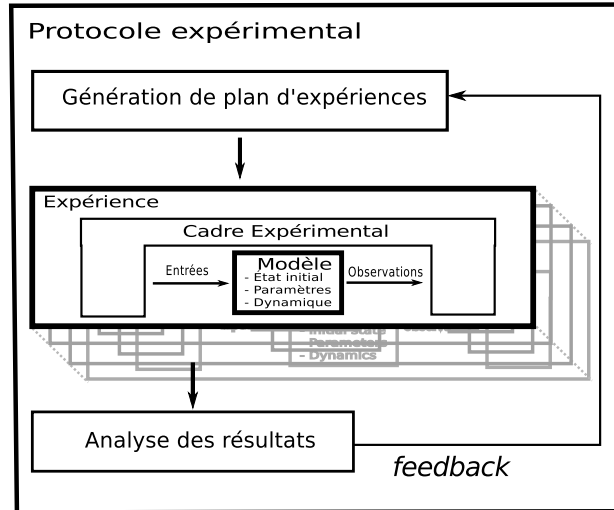


FIGURE 3.5: Protocole expérimental itératif. Il décrit comment générer les simulations à réaliser et comment analyser leurs résultats.

3.3.3 Modélisation du processus expérimental avec DEVS

Dans cette section, nous proposons une spécification du MPE utilisant le formalisme DEVS qui a été présentée dans la section 2.7.4 de l'état de l'art. Rappelons que le modèle expérimenté (A^{**} dans notre cas), tout comme le MPE, doit être spécifié en DEVS.

Le modèle de processus expérimental

L'intérêt du MPE est de pouvoir être couplé à d'autres modèles DEVS. Cela nécessite de spécifier la dynamique du protocole d'expérimentation itératif présenté dans la section 3.3.2. Par souci de modularité, le MPE est spécifié sous la forme d'un modèle couplé. Il est composé de trois modèles composants connectés comme présenté figure 3.6. Les prochains paragraphes décrivent ces trois modèles. Ces trois modèles permettent de modéliser l'activité d'un protocole expérimental tel qu'il est illustré par la figure 3.5. Le modèle de générateurs de plan d'expérience génère les plan d'expériences, le modèle expérimenteur les exécute, et le modèle analyseur de résultats traite les résultats d'expériences.

Le modèle *générateur_plan_exp* a deux ports d'entrée. Le premier reçoit des événements des modèles couplés au modèle *processus_expérimental*. Le second reçoit des événements du modèle *analyseur_res*. En utilisant les informations contenues dans ces événements, il doit comporter un ensemble de fonctionnalités lui permettant de construire un plan d'expérience tel qu'il est décrit dans la section 3.3.2 et de l'envoyer sous la forme d'un événement sur son port de sortie *plan_exp*.

Le modèle *expérimenteur* reçoit le plan d'expérience sur son port d'entrée. Le plan d'expérience est une structure de données qui contient toutes les informations nécessaires pour réaliser les simulations. Toutes les simulations sont réalisées et les résultats d'expériences sont enregistrés dans une nouvelle structure de données que l'on appelle « résultats d'expériences ». Cette structure de données est ensuite

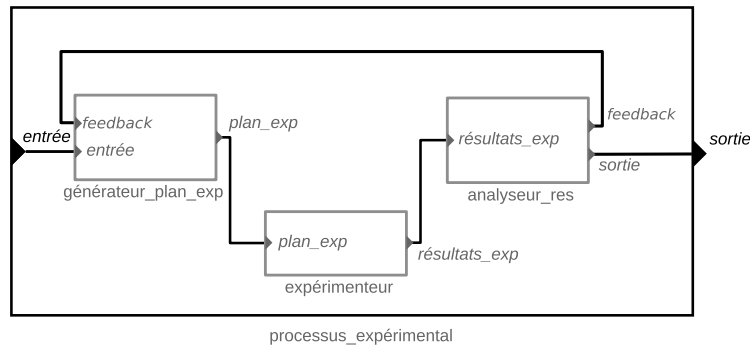


FIGURE 3.6: Structure du modèle de processus expérimental. (Image généré par le logiciel VLE.)

envoyée sur son port de sortie.

Le modèle *analyseur_res* reçoit la structure de données de résultats d'expériences et calcule à la fois les mesures de sorties à envoyer à l'extérieur du modèle *processus_expérimental*, et les mesures de *feedback* à renvoyer au modèle *générateur_plan_exp* pour éventuellement déclencher la génération d'un nouveau plan d'expérience.

Le modèle de générateur de plans d'expérience et le modèle d'analyseur de résultats sont tous deux modèle-dépendant. Pour cette raison, seuls leurs rôles et leurs caractéristiques principales peuvent être données ici. Un exemple est donné dans ce chapitre pour le cas d'un modèle de processus expérimental visant à estimer les paramètres du modèle encapsulé (cf. section 3.3.4). Un autre exemple peut être trouvé en annexe pour le cas d'un MPE conçu pour le transfert d'échelles dans un modèle multi-échelles (cf. annexe F). À l'opposé, le modèle expérimentateur est générique et peut être entièrement spécifié ici.

Spécification DEVS du modèle expérimentateur

Dans un premier temps, nous devons définir deux structures de données et une fonction.

- *PE* (Plan d'Expérience) : Une structure de données spécifiant une liste d'expériences. Ces expériences contiennent chacune les valeurs des variables suivantes :
 - la durée de simulation,
 - le graphe de connexion DEVS du modèle et du cadre expérimental,
 - les dynamiques de chaque modèle atomique,
 - les valeurs de paramètres de chaque modèle atomique,
 - les ports d'observation observés et la fréquence des observations pour chaque modèle DEVS atomique
 - le nombre de répliquas à effectuer
 - la graine du générateur de nombres aléatoires pour le premier répliqua (les graines des répliquas suivants devant être générées par le générateur aléatoire).

- *RE* (Résultats d'Expériences) : Une structure de données contenant les valeurs de sortie d'un plan d'expérience, organisés de la manière suivante :
 - le sous modèle DEVS et le port d'où l'observation provient,
 - la date de l'observation,
 - un index spécifiant l'expérience,
 - un index spécifiant le répliqua.
- *Eval* : $PE \mapsto RE$: La fonction d'évaluation utilise un moteur de simulation DEVS pour réaliser l'ensemble des simulations décrites dans une structure de donnée *PE* et remplit une structure de données *RE*. Les détails concernant une possible implémentation sont donnés dans la section « *implementation* » de l'article donné dans l'annexe F.

Nous pouvons maintenant spécifier le modèle expérimenteur comme une structure DEVS $\langle X, Y, S, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda, ta \rangle$. Son comportement dynamique est illustré par un diagramme d'états DEVS figure 3.7, et spécifié comme suit :

- *S*, l'ensemble d'états est le suivant :

$$S = \{(RE, phase, T_0) | phase \in \{INIT, PAUSE, EXP\}\}$$
 Où *RE* est une structure de donnée de résultats d'expériences telle qu'elle a été décrite et $T_0 \in \mathbb{R}$ est un temps.
- *X*, l'ensemble des évènements d'entrée est l'ensemble des paires $\{T_i, PE\}$, tels que $T_i \in \mathbb{R}$ est un temps et *PE* est une structure de données de plan d'expérience.
- *Y*, l'ensemble des évènements de sortie est l'ensemble des structures de données de résultats d'expériences possibles *RE*.
- δ_{ext} , la fonction de transition externe est la suivante :

$$\delta_{ext} : Q \times X \rightarrow S, \text{ avec } Q = \{(s, e) | s \in S, 0 \leq e \leq ta(s)\}$$

$$(PE, T_i), (RE, PAUSE, T_0) \mapsto (eval(PE), EXP, T_i)$$

Où *eval*() est la fonction d'évaluation qui calcule les résultats d'expériences en fonction du plan d'expérience.

- *ta*, la fonction d'avancement du temps est la suivante :

$$ta : S \rightarrow \mathbb{R}_0^+ \cup \infty$$

$$(RE, INIT, T_0) \mapsto 0$$

$$(RE, PAUSE, T_0) \mapsto \infty$$

$$(RE, EXP, T_0) \mapsto T_0$$

- δ_{int} , la fonction de transition interne est la suivante :

$$\delta_{int} : S \rightarrow S$$

$$\forall (RE, phase, T_0) \in S$$

$$(RE, phase, T_0) \mapsto (RE, PAUSE, T_0)$$

- λ , la fonction de sortie est :

$$\lambda : S \rightarrow Y$$

$$(RE, EXP, T_0) \mapsto RE$$

$$\forall phase \neq EXP, (RE, phase, T_0) \mapsto \emptyset$$

- δ_{con} , la fonction de transition confluyente est :

$$\delta_{con}(s, x) = \delta_{int}(\delta_{ext}(s, ta(s), x))$$

Comme nous l'avons dit précédemment, le modèle expérimenteur est générique et peut être inclus tel quel dans tout modèle de processus expérimental. Une fois implémenté, il suffit de le connecter et toutes les simulations encapsulées seront gérées par son intermédiaire. Notons certaines propriétés importantes du modèle expérimenteur. Tout d'abord, le modèle expérimenteur n'a pas de comportement autonome et va réaliser des expériences uniquement si il est perturbé par un événement extérieur. Ensuite, lors d'une telle perturbation, le modèle expérimenteur reçoit deux informations : la structure de données de plan d'expérience PE , et la durée $T_0 \geq 0$ de la phase d'expérimentation EXP . Ainsi, le fait que la réalisation d'un plan d'expérience peut éventuellement ne pas être instantanée peut être représenté. Rappelons le cas d'agents anticipatifs. Les agents peuvent avoir besoin d'un « temps de réflexion » pour réaliser des projections « dans leur tête ». La durée de ce temps de réflexion peut être spécifié par T_0 .

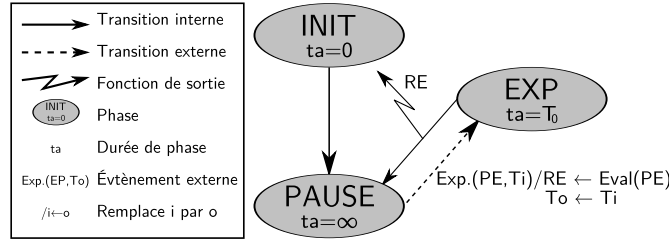


FIGURE 3.7: Diagramme d'états DEVS du modèle exécutif du système cible contrôlé.

3.3.4 Exemples d'utilisation d'un MPE

Estimation de paramètres par simulation

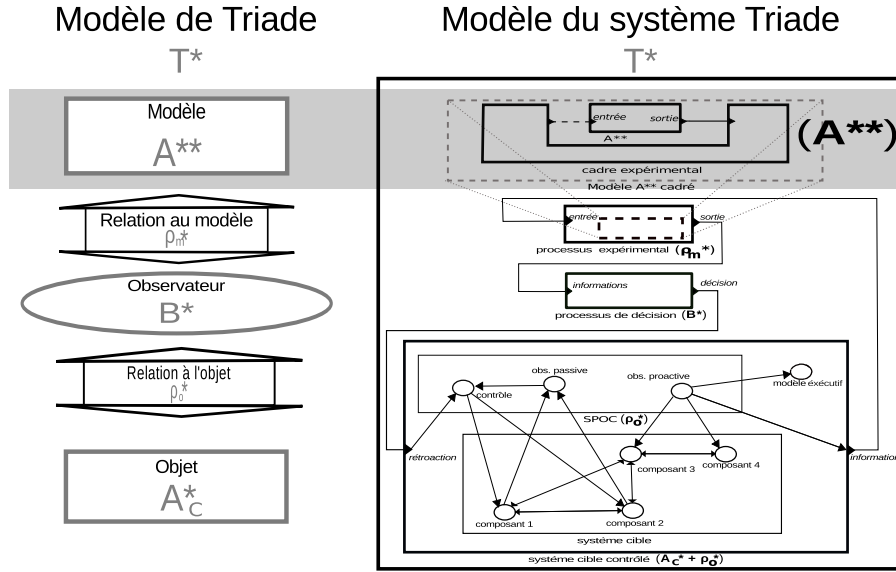
Dans le cadre de la modélisation de la triade de Minsky, une utilisation possible du modèle A^{**} est de le paramétrer de manière à ce qu'il reproduise les comportements observés chez le système cible. La méthode d'estimation de paramètres est présentée dans le chapitre précédent (cf. section 2.5.4). Un modèle de processus expérimental dédié à une méthode globale d'estimation de paramètres doit être capable de construire la série temporelle $(y(t_i))_{i=1..N}$ au fur et à mesure de la réception des observations. À tout instant t_N , il doit pouvoir estimer le jeu de paramètres θ^* tel qu'une mesure de distance $D(\theta^*, (y(t_i))_{i=1..N})$ entre les données et les résultats du modèle pour le jeu de paramètres à estimer, soit minimum. Cette minimisation est assurée par un algorithme d'optimisation par simulation.

Nous présentons un MPE d'estimation de paramètres basé sur une méthode d'optimisation par essaim de particules dans l'annexe D.

Modélisation multi-échelles

Outre l'article présenté en annexe (cf. annexe F), d'autres travaux avaient été réalisés antérieurement sur l'utilisation d'un MPE pour la modélisation multi-échelles [BDQM09]. Ces premiers travaux proposaient une autre formalisation du MPE qui utilisait une décomposition hiérarchique différente. Le MPE était constitué d'un seul système au lieu de trois sous-systèmes couplés.

3.4 A^{**} : Le modèle A^{**} cadré



Le modèle A^{**} est un MSD spécifié avec le formalisme DEVS il est simulé en même temps qu'un cadre expérimental qui est également un MSD. La figure 3.8 donne la représentation classique d'un cadre expérimental couplé à un modèle. Nous renvoyons le lecteur à la section 2.7.2 (page 42) pour une description plus précise des cadres expérimentaux, à l'annexe D pour un exemple de cadre expérimental, et au chapitre d'application pour un exemple de modèle A^{**} cadré dans un modèle de triade de Minsky (cf. section 4.4.3 page 89).

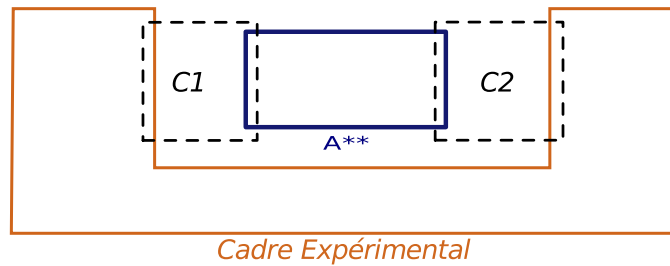
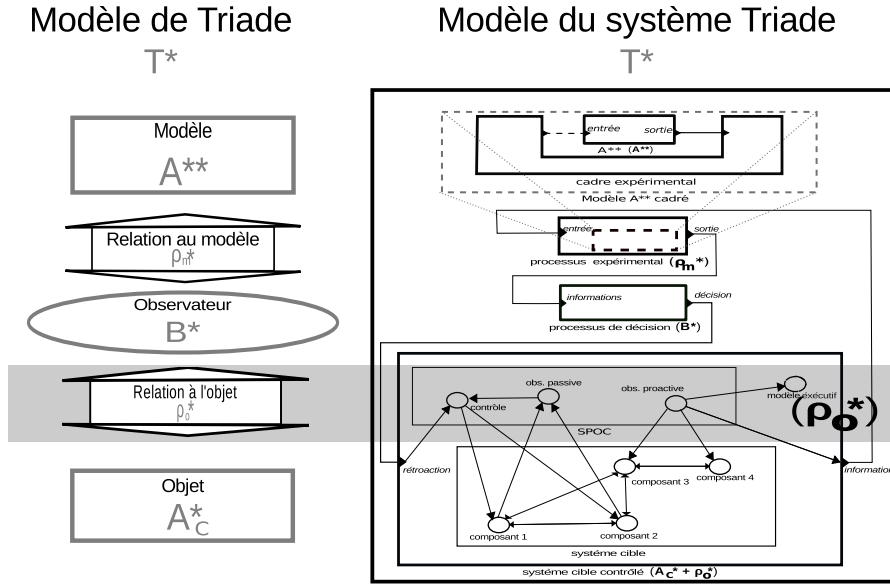


FIGURE 3.8: Modèle A^{**} cadré. Le cadre C1 représente les connexions entre les ports de sorties du cadre expérimental et les ports d'entrée du modèle A^{**} . Le cadre C2 représente les connexions entre les ports de sorties du modèle A^{**} et les ports d'entrée du cadre expérimental.

3.5 ρ_o^* : Le modèle de SPOC



Le Système Pour l'Observation et le Contrôle incarne la relation à l'objet ρ_o de l'observateur de Minsky. Il regroupe l'ensemble des processus de contrôle et d'observation attribués à l'observateur de Minsky dans la Triade de Minsky. Il est présenté dans la figure ?? comme un ensemble de sous-systèmes composants du système cible contrôlé. En effet, les processus d'observation peuvent être très dépendants du système cible. Par exemple, un processus d'observation peut n'observer qu'une partie du système cible. De plus, les processus d'observation peuvent interagir avec des processus de contrôle du système cible, qui peuvent être « localisés » également. En conséquence, nous ne proposons pas de modèle de SPOC mais un ensemble de modèles de processus d'observation et de contrôle. Rappelons que nous considérons des cas où le modèle de système cible est formalisé sous la forme d'un modèle couplé (DS-DEVS). Dans ce cas, les modèles de processus d'observation et de contrôle que nous avons développés doivent être intégrés dans le modèle de système cible contrôlé pour y simuler l'activité du SPOC. Le fait de les situer au même niveau hiérarchique que les composants du système cible permet de clarifier la relation étroite qu'ils entretiennent avec le système cible et surtout de faciliter grandement l'implémentation, puisque toute la gestion des connexions entre les modèles de processus d'observation et de contrôle d'une part, et les sous-modèles composants du modèle de système cible d'autre part, se fait via l'intermédiaire du seul modèle exécutif du modèle de système cible contrôlé. Nous verrons dans la suite les contraintes que doivent respecter le modèle exécutif et les modèles composants du modèle de système cible. Un exemple de modèle exécutif de système cible contrôlé est donné section 3.5.3.

3.5.1 Les modèles du système d'observation

Le système d'observation regroupe l'ensemble des processus qui récoltent de l'information brute sur le système cible pour l'observateur de Minsky. Il s'agit d'un système qui reçoit des informations sur l'état du système cible, les stocke, les traite et les envoie au système de processus expérimental. Il peut en outre envoyer lui-même des requêtes au système cible et récupérer les résultats.

Nous avons distingué trois types de processus d'observation dont la combinaison doit nous permettre de décrire toute stratégie d'observation. Le premier est le processus d'observation passive qui consiste à simplement recevoir, stocker et éventuellement transmettre des événements contenant de l'information. Le second est le processus d'observation proactive qui observe régulièrement certaines parties du système à des dates données. Enfin le troisième est le processus d'observation réactive qui consiste à observer le système en réaction à un événement extérieur qui identifie éventuellement des parties du système cible à observer. (L'événement extérieur peut être par exemple une observation résultant d'un processus d'observation passive.)

Pour chacun de ces processus, un modèle DEVS a été implémenté qui pourra être paramétré et introduit dans le modèle de système cible contrôlé sous la forme d'un sous-modèle composant. Il s'agit des modèles d'observation passive, proactive et réactive. Un modèle supplémentaire de centralisation d'informations a également été conçu pour centraliser les informations provenant des différents processus d'observation. Décrivons ici leurs principales caractéristiques et les contraintes que leur utilisation implique sur le modèle de système cible. Par souci de lisibilité, nous utilisons dans la suite le terme de « modèle composant » pour désigner les modèles composants du modèle de système cible qui ne sont ni des modèles du SPOC, ni le modèle exécutif du modèle de système cible. De la même manière nous utilisons le terme modèle exécutif pour parler du modèle exécutif du modèle de système cible contrôlé.

Le modèle d'observation passive

Il dispose d'un port d'entrée « status » et d'un port de sortie « centre d'information ». Si un modèle d'observation passive est utilisé, les modèles composants qui sont susceptibles d'être observés par ce processus doivent disposer d'un port de sortie « status » sur lequel ils envoient à chaque transition d'état, un événement contenant la valeur de l'état observé et leur identifiant au sein du modèle de système cible contrôlé. À la réception d'un événement, le modèle d'observation passive effectue un traitement de l'information et un test de transmission. Le traitement de l'information consiste en une modification éventuelle de la valeur de l'état ou de l'identifiant du composant de système cible. Le test de transmission détermine si l'information (éventuellement modifiée) est transmise sous la forme d'un événement de sortie envoyé sur son port de sortie. Un délai peut être paramétré entre le moment où l'information est reçue et celui où elle est transmise. Ce délai est implémenté sous la forme d'un modèle de délai à coupler au modèle d'observation passive (cf. annexe C.1). Le modèle exécutif doit être en mesure d'établir la connexion entre les modèles composants et le modèle d'observation passive (cf. section 3.5.3).

Le schéma de connexion d'un modèle d'observation passive est présenté figure 3.9. La dynamique d'un modèle d'observation passive est résumée par le diagramme d'états présenté figure 3.10.

Le modèle d'observation proactive

Il effectue des observations à pas de temps constant (paramétré). Ce modèle a un port d'entrée « status » et deux ports de sortie. Un port « connecte » doit être relié au modèle exécutif, et un port « status? » doit être connecté, à une date

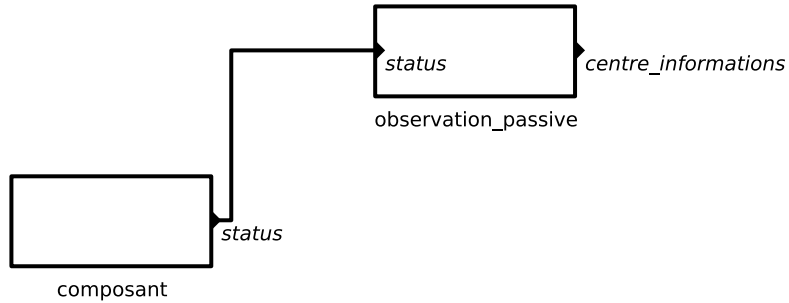


FIGURE 3.9: Schéma de connexion entre un modèle d'observation passive et un modèle composant du système cible.

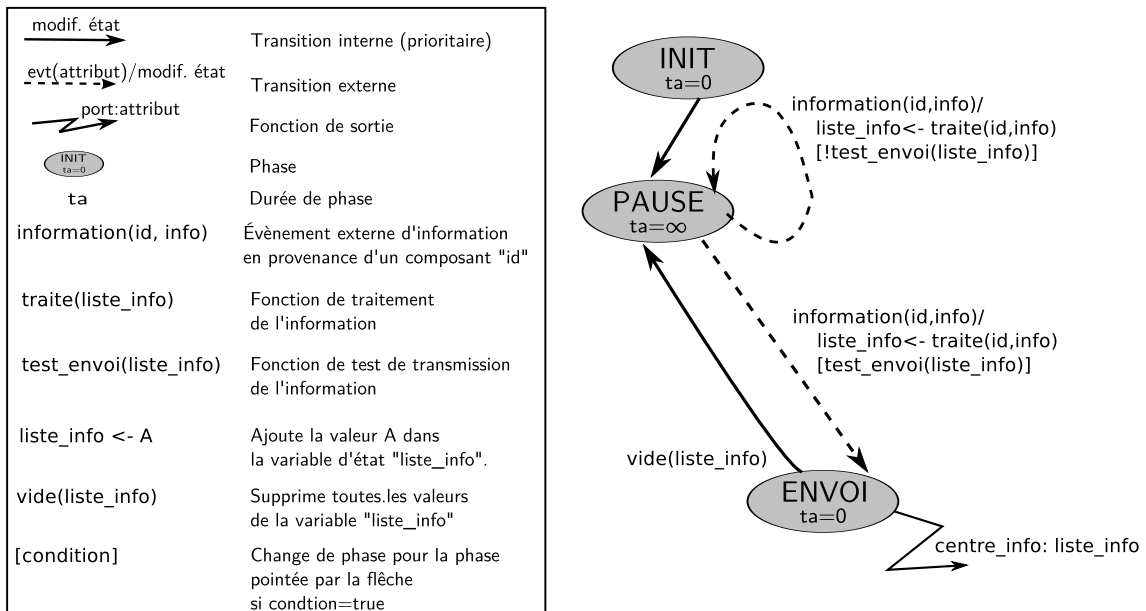


FIGURE 3.10: Diagramme de phase DEVS d'un modèle d'observation passive

d'observation donnée, à l'ensemble des modèles composants observés à cette date. À chaque pas de temps, le modèle d'observation proactive commence par générer la liste des modèles composants à observer. (Il doit pour cela disposer de la liste des identifiants des modèles composants du système cible et d'une fonction de choix des composants à observer). Il envoie ensuite un message de connexion sur son port « connecte ». Ce message contient la liste des identifiants des composants à observer et son propre identifiant dans le modèle de système cible. À la réception de cet événement, le modèle exécutif doit détruire toutes les connexions existant entre le modèle d'observation proactive et les modèles composants, puis il doit relier le modèle d'observation proactive à l'ensemble des composants dont l'identifiant figure dans la liste du message de connexion. Le schéma de connexion entre un modèle d'observation proactive et un modèle composant est présenté figure ?? . Un port de sortie ayant pour nom l'identifiant du modèle d'observation proactive est créé si il n'existait pas déjà. Le modèle d'observation proactive envoie ensuite un message de demande d'information sur son port « status? ». Ce message d'information contient l'identifiant du modèle d'observation proactive. Chaque modèle composant recevant ce message renvoie immédiatement son état sur son port de sortie ayant le nom de l'identifiant du modèle d'observation proactive. À la réception d'un événement sur son port « status », le modèle d'observation proactive traite l'information par une fonction de traitement de l'information puis la stocke. Après un temps très petit qui assure que tous les événements ont été reçus pour une phase d'observation, l'information est agrégée par une fonction d'agrégation et envoyée sur le port « centre d'information ».

La dynamique d'un modèle d'observation proactive est résumée par le diagramme d'états présenté figure 3.12.

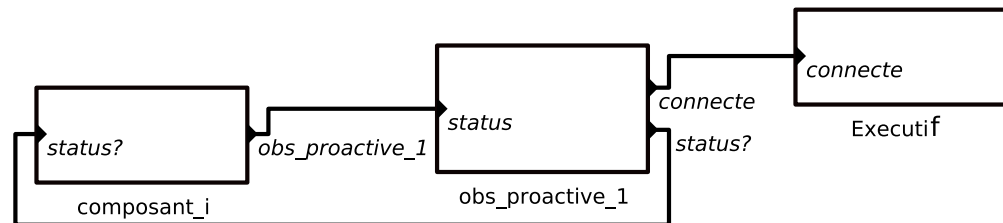


FIGURE 3.11: Schéma de connexion entre un modèle d'observation proactive et un modèle composant du système cible.

Le modèle d'observation réactive

Il fonctionne de la même manière que le modèle d'observation proactive, à la différence qu'il ne génère pas les dates d'observation, et se base sur des événements externes pour générer la liste des composants à observer. Le protocole de récolte d'information est identique à celui décrit pour le modèle d'observation proactive à un pas de temps d'observation. La différence est qu'il est enclenché à la réception d'un événement extérieur sur un port « nouvelle observation » et que la liste des

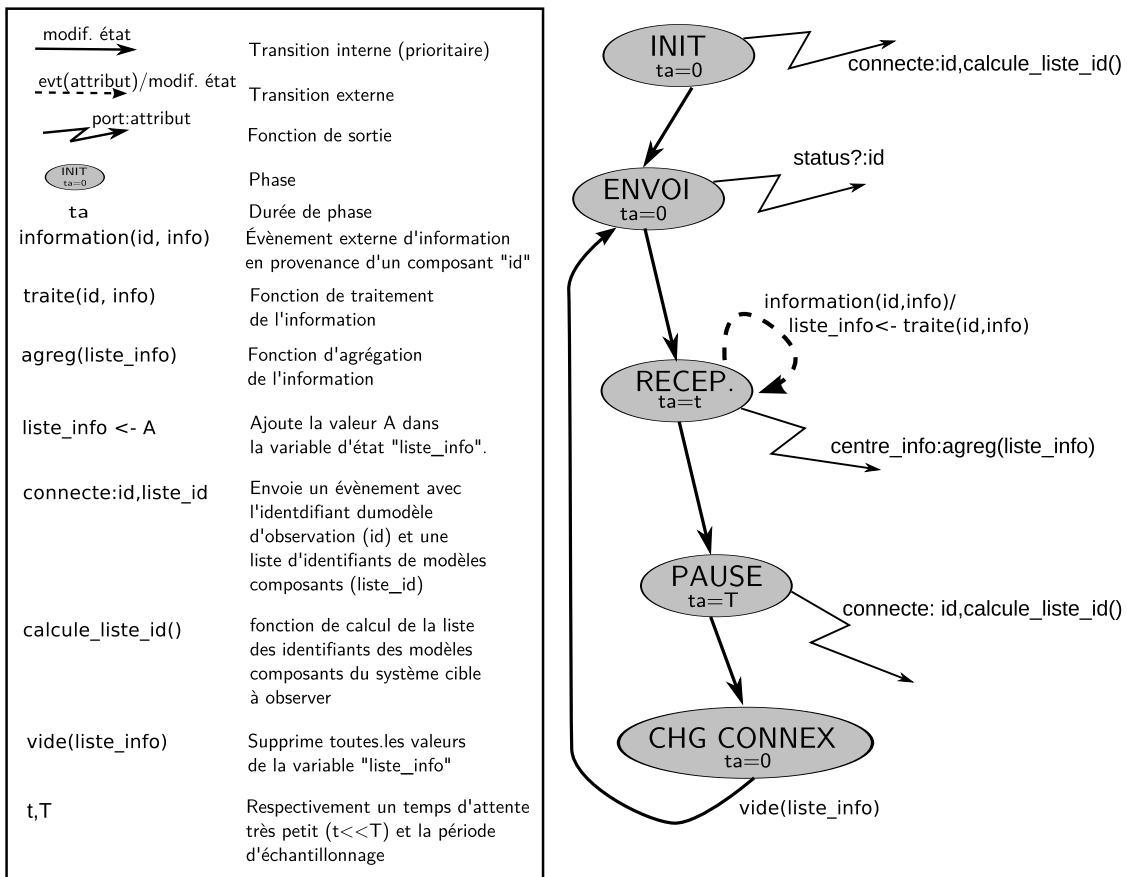


FIGURE 3.12: Diagramme de phase DEVS d'un modèle d'observation proactive

composants à observer est générée à ce moment là en fonction des attributs de cet évènement.

La figure 3.13 présente le schéma de connexion d'un modèle d'observation réactive. La dynamique d'un modèle d'observation réactive est résumée par le diagramme d'états présenté figure 3.14.

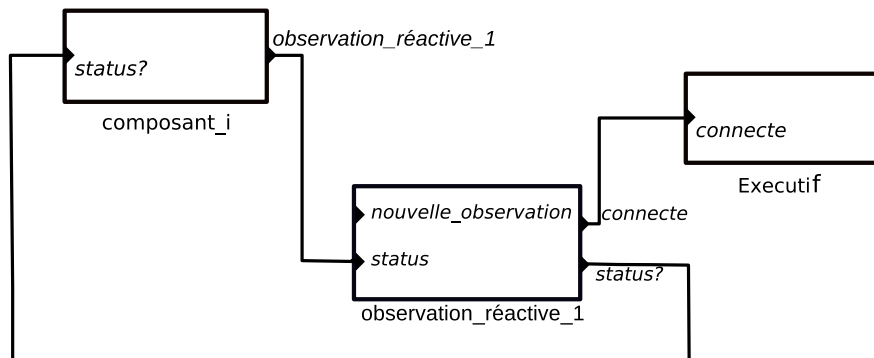


FIGURE 3.13: Schéma de connexion entre un modèle d'observation réactive et un modèle composant du système cible.

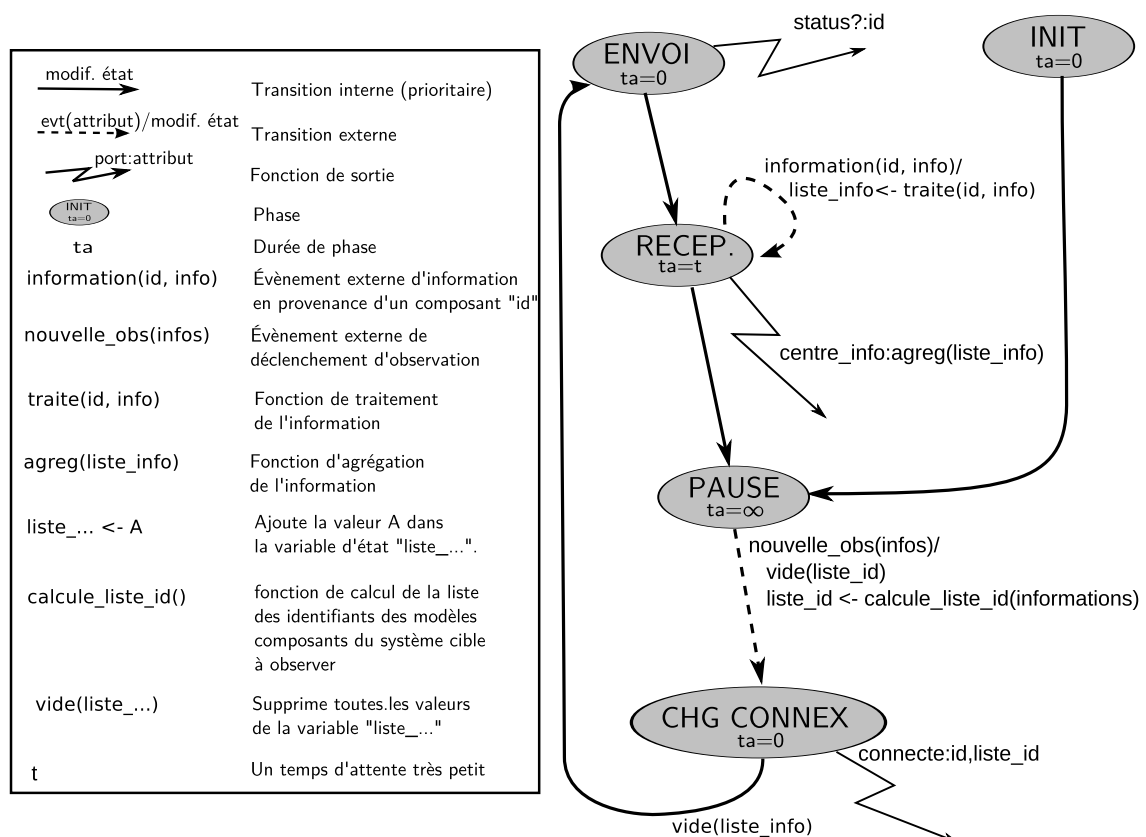


FIGURE 3.14: Diagramme de phase DEVS d'un modèle d'observation réactive

Notons que le modèle d'observation réactive peut être utilisé pour spécifier un planning d'observation prédéfini. Il doit pour cela être couplé à un modèle ordon-

nanceur. Le modèle ordonnanceur doit être paramétré par un calendrier qui liste les observations à effectuer et les dates de chaque observation. À une date d'observation apparaissant dans son calendrier, le modèle ordonnanceur envoie un message de nouvelle observation au modèle d'observation réactive. Ce message contient la liste des identifiants des composants à observer à cette date. Appelons le modèle couplé composé d'un modèle ordonnanceur et d'un modèle d'observation proactive, un modèle d'observation planifiée. Il est présenté en annexe (cf. annexe C.2).

Le tableau 3.1 présente la composition des événements utilisés dans le protocole de communication entre le modèle de système cible et le modèle de SPOC.

3.5.2 Les modèles du système de contrôle

Tout comme le système d'observation le système de contrôle est composé de processus en relation étroite avec le système cible. Les processus de contrôle reçoivent des informations provenant de processus d'observation et peuvent effectuer des actions de contrôle sur certaines parties du système cible. La politique de contrôle peut être modifiée par le système de décision. Nous ne proposons pas de modèles de processus de contrôles génériques mais nous pouvons les situer dans le modèle de système triade en disant qu'ils doivent être placés à l'intérieur du modèle de système cible contrôlé, être connectés en entrée à des modèles de processus d'observation et/ou au modèle de processus de décision et en sortie à des composants du modèle de système cible et éventuellement au modèle de centralisation d'information du modèle de SPOC.

3.5.3 Le modèle de système cible contrôlé

Les événements de contrôle ou d'observation envoyés par les modèles de processus de contrôle ou d'observation doivent être formatés comme présentés dans le tableau 3.1. De la même manière que pour les événements de demande d'observation, les sous-modèles composants du modèle de système cible doivent être en mesure de réagir aux événements de contrôle reçus.

Modèle exécutif

Rappelons que le modèle exécutif contient les autres modèles composants de son niveau hiérarchique dans son état (selon la spécification DSDE rappelée section 2.7.4 page 48). Il n'a donc pas besoin d'être « connecté » aux modèles composants par leurs ports d'entrée/sortie pour pouvoir modifier les connexions du modèle composé auquel il appartient. En revanche, si un modèle composant veut avoir accès à la structure de ce niveau hiérarchique, il doit le faire par l'intermédiaire du modèle exécutif en lui envoyant un événement. Le modèle exécutif du modèle de système cible contrôlé possède donc un port d'entrée à cet effet, le port « connecte » (cf. figure 3.15).

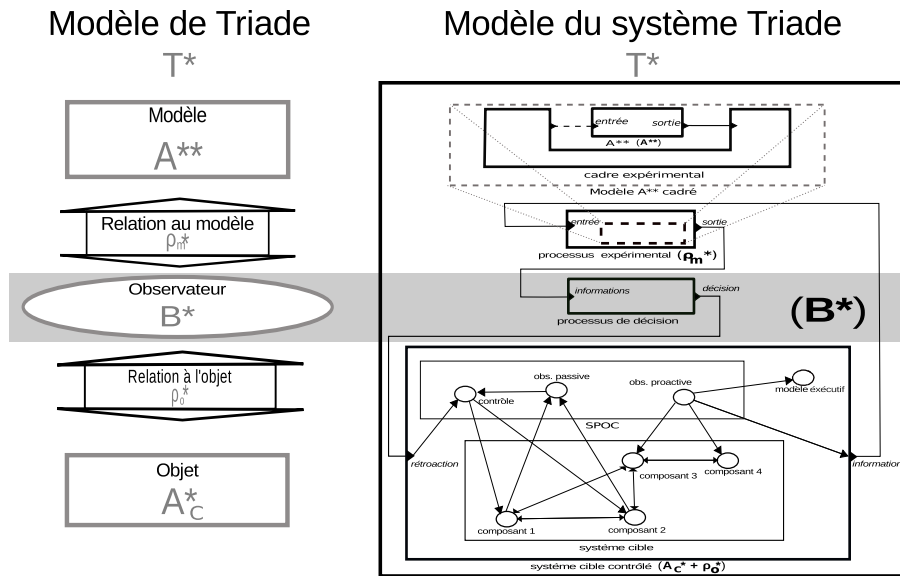
La spécification formelle du modèle exécutif est une structure P-DEVS classique : $M_X = \langle X, Y, S, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda, ta \rangle$. Un modèle exécutif de système cible contrôlé est présenté ici à titre d'exemple. À l'initialisation de la simulation, il connecte tous les modèles de contrôle et d'observation passive à tous les modèles composants du système cible. Ensuite, à chaque réception d'événement de changement de connexion

Origine (port)	Dest. (port)	Attributs (type)
comp(status)	obsPass(status)	id(string), état(libre)
obsAct(status?), obsRéact(status?)	comp(status?)	id(string)
comp(id_dest)	obsAct(status), obsRéact(status)	id(string), état(libre)
obsAct(connecte), obsRéact(connecte)	exe(connecte)	id(string), liste_id(liste de string)
contr(id_dest)	comp(contr)	controle_type(string), attributes(libre)

TABLE 3.1: Évènements échangés dans le cadre du protocole de communication entre le modèle de SPOC et le modèle de système cible. Les abréviations obsPass., obsAct., obsRéact., contr., comp. et exec., désignent respectivement des modèles d'observation passive, active, et réactives, des modèles de contrôle, des modèles composants et le modèle exécutif.

(envoyé par les modèles d'observation active), il modifie les connexions de ces modèles. Sa dynamique est résumée par le diagramme d'états présenté figure 3.16.

3.6 B* : Le modèle de processus de décision



Le processus de décision représente la traduction de l'information générée par le processus expérimental en une modification du SPOC. Il peut faire intervenir des règles de décision multiples et dépendantes à la fois du type d'informations fournies par le processus expérimental et du type d'actions possibles déterminées par le SPOC. Il est donc difficile d'en donner une structure générique. Ce que l'on peut en dire à ce stade est qu'on le considère comme un système qui reçoit les informations produites par le processus expérimental et les utilise pour intervenir sur les stratégies de contrôle et d'observation. Nous l'appellons le processus de décision. C'est par cette rétroaction que l'activité de simulation, en approvisionnant les décisions prises sur

le SPOC, a un impact sur le système cible.

Dans ce cadre générique, nous nous contentons de placer le modèle de système de décision dans le modèle T^* .

3.7 Implémentation et expérimentation de T^* avec VLE

Pour pouvoir expérimenter le modèle de système triade que nous avons décrit tout au long de ce chapitre, il est nécessaire d'implémenter un simulateur qui va nous permettre de le simuler. L'activité qui consiste à expérimenter le modèle T^* correspond à la relation ρ_{m_c} de la triade T' (cf. figure 3.1).

Les composants du modèle T^* présentés dans cette partie ont été implémentés en utilisant le Virtual Laboratory Environment (VLE) [QDR09], un ensemble d'outils et d'interfaces de programmation basé sur DEVS pour la modélisation et la simulation des systèmes complexes. Les structures de plan d'expériences et de résultats d'expériences (PE et RE) sont implémentées en utilisant respectivement les représentations utilisées par VLE pour les plans d'expériences et les résultats d'expériences. Ces structures sont sérialisables, c'est à dire qu'elles peuvent être incluses dans l'état d'un modèle et sérialisées, puis désérialisées de manière à être transmises par des événements. De plus le VLE permet la simulation encapsulée. C'est à dire que l'interface de programmation d'un modèle DEVS contient des fonctions qui permettent de lancer des simulations en utilisant le moteur de simulation de VLE. Nous avons utilisé ces fonctionnalités pour la programmation du modèle expérimenteur du MPE .

VLE dispose en outre d'une interface avec le logiciel R dédié à l'analyse statistique [R D06] qui permet de paramétrer des expériences, de lancer des simulations et de récupérer les résultats depuis une console R. Ainsi, les packages R dédiés à l'exploration de l'espace des paramètres ou à l'analyse de sensibilité peuvent être utilisés pour expérimenter un modèle T^* implémenté en VLE.

La figure 1.8 positionne les développements informatiques de cette thèse au niveau des entités T^* et ρ_{m_c} de la triade T' . Rappelons que l'entité T^* est composée de A_C^* , ρ_o^* , B^* , ρ_m^* et A^{**} qui correspondent à des modèles des entités de T , respectivement au modèle de l'objet, au modèle de la relation à l'objet, au modèle de l'observateur de Minsky, au modèle de la relation au modèle et au modèle du modèle. Le tableau 3.2 résume comment définir et implémenter ces entités.

Entité de T'	Formalismes	Outils conceptuels	Outils informatiques disponibles
A^{**} cadré	DEVS, formalismes conformes à DEVS, Multiformalismes conformes à DEVS	Cadre Expérimental	VLE, modèles générateur et accepteur du CE_C (cf. annexe D)
ρ_m^*	DEVS	MPE	VLE, modèle expérimenteur du MPE , structure du MPE et protocole de communication entre les entités du MPE , modèles MPE_{OEP} (cf. annexe D)
ρ_o^*	DS-DEVS, formalismes conformes à DEVS, Multiformalismes conformes à DEVS	SPOC et proposition d'intégration du modèle de SPOC au modèle de système cible	VLE, modèles de processus d'observation, protocole d'interaction avec A_C^*
B^*	DEVS, Théorie de la décision (cf. section 2.6.1)	position dans le modèle de système triade	
A_C^*	DEVS, formalismes conformes à DEVS, Multiformalismes conformes à DEVS	Proposition d'intégration du SPOC à A_C^*	VLE, protocole d'interaction avec ρ_o^*
ρ_{m_c}	Libre	Méthodes statistiques d'analyse de sensibilités, d'optimisation d'exploration d'espace des paramètres, ...	interface entre le logiciel R et VLE

TABLE 3.2: Définition, implémentation et expérimentation de T*

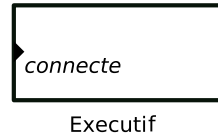


FIGURE 3.15: Modèle exécutif du modèle de système cible contrôlé.

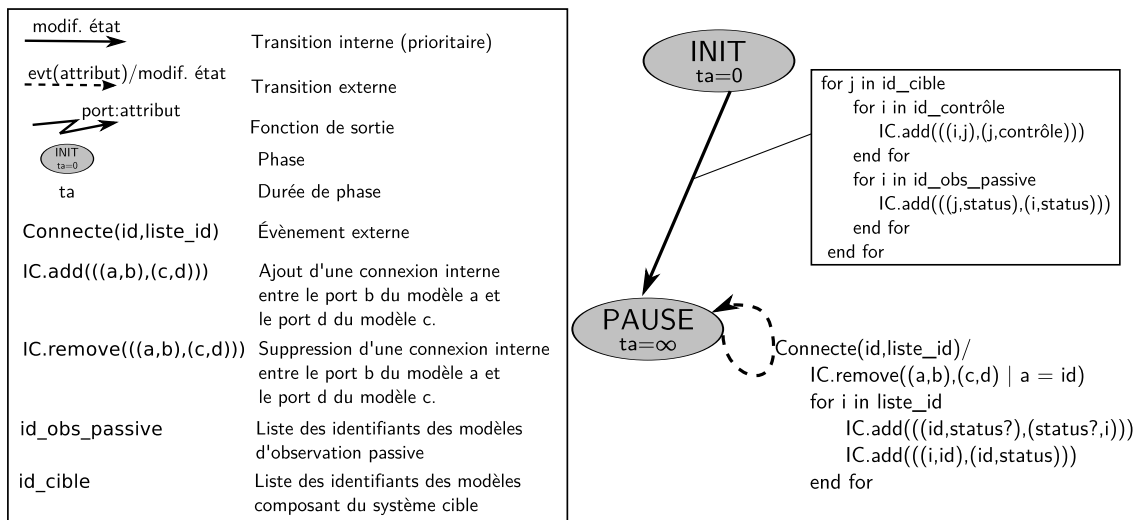


FIGURE 3.16: Diagramme de phase DEVS d'un modèle exécutif de système cible contrôlé

Chapitre 4

Application à la surveillance épidémiologique en santé animale

RÉSUMÉ

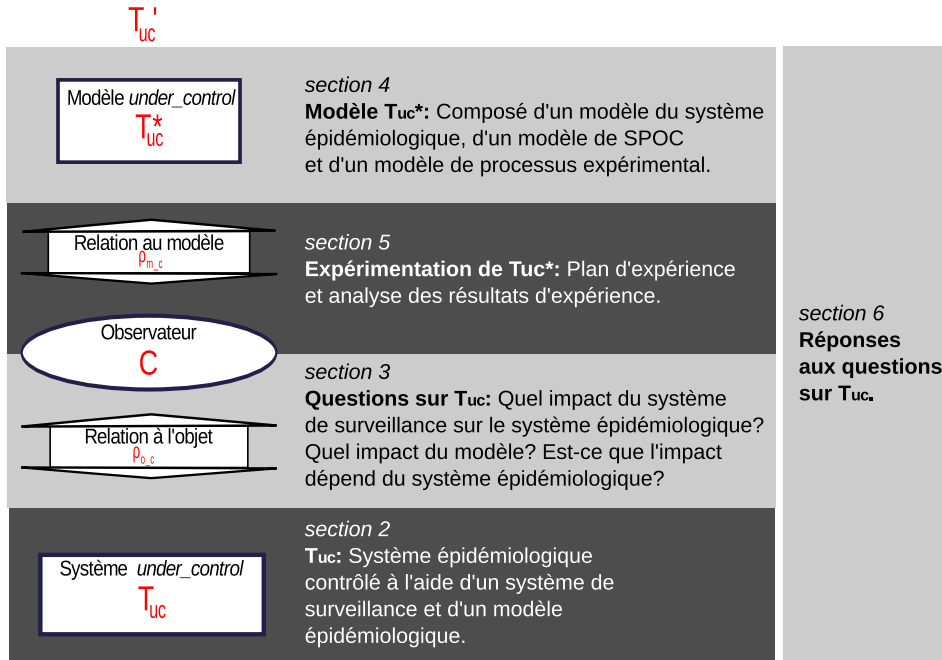
Notre application est un exemple didactique de triade T' en épidémiologie que l'on note T_{uc}' . L'objet de Minsky de T_{uc}' est une triade de Minsky notée T_{uc} (UC pour *Under Control*) dans laquelle : le système cible est un système épidémiologique (Un ensemble d'unités épidémiologiques reliées par un réseau de contacts infectieux) ; le SPOC comprend un système de contrôle de la maladie et un système d'épidémiosurveillance mis en place pour évaluer les actions du système de contrôle ; le processus expérimental consiste à calibrer un modèle épidémiologique (modèle à compartiment SIS) sur les données de surveillance en vue de calculer l'indicateur R_0 de la maladie pour un niveau de contrôle donné. Un ensemble de questions de recherches sont identifiées au sujet de la triade T_{uc} . Un modèle expérimentable T_{uc}^* est construit pour répondre à ces questions. Un plan d'expérience est élaboré, les expériences sur T_{uc}^* (des simulations) sont réalisées et les résultats d'expériences sont analysés. Enfin, les enseignements sur le système T_{uc} et des questions supplémentaires qui pourraient être traitées avec le modèle T_{uc}^* sont discutés.

4.1 Introduction : La triade T_{uc}'

Ce chapitre présente l'application de notre cadre à la surveillance épidémiologique en santé animale. Il décrit une instance de la triade T' comme illustrée figure 1.8 page 22. Nous appellerons cette triade T_{uc}' car son objet d'étude est la triade de Minsky T_{uc} que nous décrivons ci-après. La question de l'observateur de Minsky de T_{uc}' porte sur les processus d'observation et d'utilisation de modèles à l'intérieur de T_{uc} , et le modèle de Minsky de T_{uc}' (noté T_{uc}^*) est formalisé et expérimenté en utilisant le cadre opérationnel proposé dans le chapitre précédent (chapitre 3).

La première section de ce chapitre présente le système « *under control* » (T_{uc}). Il s'agit d'une situation imaginaire représentative de l'utilisation d'un système de sur-

veillance et d'un modèle épidémiologique pour le contrôle d'une maladie infectieuse. La seconde section présente un ensemble de questions de recherche identifiées au sujet de la triade T_{uc} . La troisième section présente le modèle « *under control* » (T_{uc}^*) construit pour répondre à ces questions. La quatrième section présente l'expérimentation de T_{uc}^* , c'est-à-dire le plan d'expérience réalisé et les résultats de simulation obtenus. Enfin la dernière section discute comment les résultats obtenus sur T_{uc}^* permettent de répondre aux questions posées sur T_{uc} . La figure 4.1 illustre la triade T_{uc}' et le plan de ce chapitre.

FIGURE 4.1: La triade T_{uc}'

4.2 Système « *under control* » (T_{uc})

Le système T_{uc} représente une situation épidémiologique théorique imaginée sur la base de lectures et d'un ensemble de discussions tenues avec des membres de l'équipe AGIRs. Cette situation théorique a pour but de constituer l'objet de Minsky de la triade T_{uc}' . Le tableau 4.1 donne le vocabulaire épidémiologique utilisé dans ce chapitre.

4.2.1 Situation d'épidémie contrôlée

Nous considérons une situation tirée de la surveillance épidémiologique des maladies infectieuses en santé animale. Dans cette situation, on imagine une maladie transmissible exotique risquant d'être introduite dans une région géographique contenant des élevages d'animaux susceptibles à cette maladie. La maladie n'est pas une maladie à déclaration obligatoire mais elle peut entraîner des pertes de production pour les élevages infectés. Deux mesures ont donc été prises pour empêcher que

Vocabulaire	Définition
Prévalence	Proportion de cas dans une population à un instant donné
Épidémie	Augmentation rapide du nombre de cas dans une population
Endémique	Se dit d'une maladie qui se maintient dans une population

TABLE 4.1: Petit glossaire épidémiologique

la maladie ne devienne endémique dans la région. La première est la mise en place d'un réseau d'épidémiovigilance de manière à détecter l'introduction de la maladie si elle survient. (On peut imaginer que les vétérinaires de la région concernée ont été formés à la détection de cas de la maladie et qu'un réseau de remontée d'information a été mis en place). La seconde mesure est la mise en place de mesures de contrôle des mouvements d'animaux dans la région (on pense par exemple à la mise en place de postes de contrôle sur les routes principales). En temps normal, seule la mesure d'épidémiovigilance est en place. Dès lors que la maladie est détectée, une alerte est déclenchée et les mesures de contrôle sont utilisées pour imposer une restriction des mouvements d'animaux à un niveau d'intensité par défaut noté niveau d'intensité 0. La conséquence de cette mesure de contrôle de niveau 0 est que les mouvements d'animaux sont réduits d'un ratio $x_0\%$ inconnu. On se retrouve en présence d'une épidémie contrôlée. Cependant, restreindre les mouvements d'animaux entraîne également une perte de production à l'échelle de la région. En conséquence, une réévaluation de l'intensité de la mesure de restriction des mouvements est prévue à intervalles de temps réguliers. Un réseau d'épidémiosurveillance est prévu pour permettre cette réévaluation. Ce réseau doit effectuer des enquêtes longitudinales (i.e. étalées dans le temps) qui doivent fournir une estimation de l'évolution de la prévalence dans la région. Sur la base de ces estimations de prévalence, un modèle est utilisé pour estimer si le niveau d'intensité de la restriction des mouvements est insuffisant, juste suffisant ou si il semble démesuré et peut être diminué. La figure 4.2 illustre la chronologie des mesures prises dans cette situation.

4.2.2 Système T_{uc}

Le système T_{uc} est composé d'un système cible, d'un SPOC et d'un processus expérimental. Le système cible (également appelé système épidémiologique dans la suite du chapitre) est un groupe de cent unités épidémiologiques (des sous régions de la région géographique concernée) connectées entre elles par un réseau infectieux via lequel une unité épidémiologique infectée peut infecter ses voisines. Le SPOC est composé d'un système d'observation passive, d'un système d'observation proactive, d'un système de contrôle et d'un système de décision. Le système d'observation passive observe l'ensemble des composants du système cible jusqu'à ce qu'un foyer soit détecté. Lorsqu'un foyer est détecté, les systèmes d'observation proactive et de contrôle sont activés. Le système de contrôle impose alors la mesure de restriction des mouvements d'animaux dans le système cible avec une intensité par défaut. Il peut modifier cette intensité par une autre à la réception d'un message du système

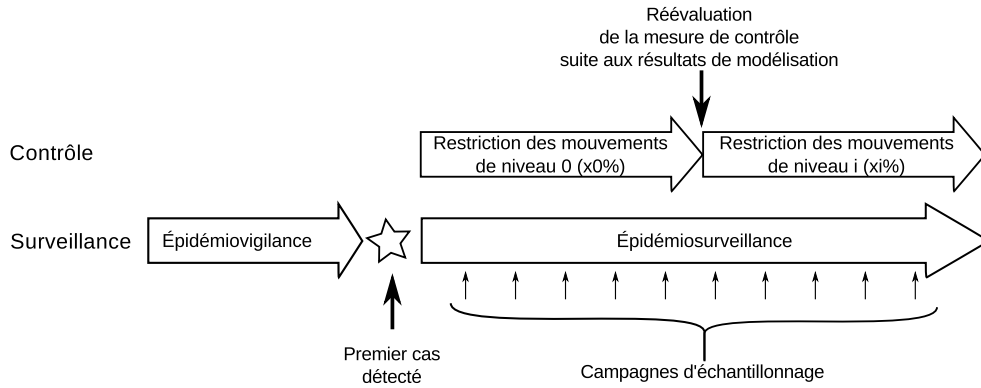


FIGURE 4.2: Chronologie des mesures de surveillance et de contrôle dans la situation d'épidémie contrôlée. Les flèches horizontales représentent des processus continus dans le temps et les flèches verticales des événements ponctuels.

de décision lui donnant la nouvelle intensité à imposer. Le système d'observation proactive observe à pas de temps régulier un échantillon représentatif d'unités épidémiologiques (qui sont les systèmes composants du système cible) de manière à calculer à chaque pas de temps une estimation de la prévalence dans le système cible. Ces estimations sont transmises au processus d'expérimentation. Le système de décision reçoit les informations du processus expérimental (il s'agit de la valeur de l'indicateur R_0) et décide d'après cette valeur, d'augmenter, de diminuer ou de ne pas changer l'intensité du contrôle. Enfin, le processus d'expérimentation consiste à calibrer un modèle compartimental de type SIS (cf. chapitre 2 section 2.3.2) dès qu'un nombre suffisant de valeurs de prévalence est disponible et à déduire de ce modèle la valeur du R_0 (cf. chapitre 2 section 2.5.1) de la maladie pour un niveau de contrôle donné.

4.3 Questions sur le système T_{uc}

Les questions que l'on se pose sur T_{uc} dans le cadre de la triade T_{uc} sont les suivantes :

- 1 Y a-t-il un impact des caractéristiques du système d'épidémiosurveillance sur les pertes de production causées par la maladie ? Si oui, lequel ? Est-il possible ou nécessaire de modifier le système de surveillance en conséquence ?
- 2 Quel impact a le choix d'un processus expérimental sur les pertes de production causées par la maladie ? Par exemple, est-ce que la connaissance du taux de rétablissement des unités épidémiologiques (qui est un des deux paramètres du modèle SIS) permettrait de réduire les pertes de production causées par la maladie ?
- 3 Est-ce que ces impacts dépendent des caractéristiques du système épidémiologique ? Si oui, est-il possible de dégager une typologie de systèmes épidémiologiques qui associe un type de système épidémiologique à des contraintes à respecter pour le système de surveillance ou pour le processus expérimental ?

Elles sont classées en trois catégories car elles concernent des entités ou des relations entre entités différentes de la triade T_{uc} . Pour expliciter ces propos, attribuons des lettres associées à la triade de Minsky aux composant du système T_{uc} . Notons A le système épidémiologique, ρ_o le SPOC décrit ci-dessus, ρ_m le processus expérimental et A_B^* le modèle SIS¹. Les questions de type 1 portent sur la conception du système d'observation proactive (relation ρ_o). Les questions de type 2 portent sur la conception du processus expérimental (relation ρ_m). Enfin, les questions de type 3 portent sur la pertinence du choix du système d'observation proactive et du processus expérimental (ρ_o et ρ_m), en fonction de la nature du système épidémiologique (A).

Ces questions portent toutes sur l'évaluation de l'impact qu'auront certains choix de conception du SPOC ou du processus expérimental sur les conséquences de la maladie sur le système cible (éventuellement en fonction de la nature du système cible pour les questions de type 3). Pour cet exemple, nous allons nous appliquer à tester un nombre restreint d'hypothèses de conception du système de surveillance et de modes d'utilisation du modèle SIS sur un certain nombre de scénarios de maladie. La description du système T_{uc} est suffisamment précise pour que nous puissions identifier des variables qui vont servir d'une part à caractériser différentes modalités de conception du système de surveillance et d'utilisation de modèles, et d'autre part à caractériser différents types de maladies. Nous appelons ces variables des « facteurs ». Un ensemble d'indicateurs sont identifiés pour évaluer les conséquences de la maladie sur le système cible et sur le système triade dans son ensemble. Nous appelons ces indicateurs des « critères » car ils vont servir à évaluer les hypothèses de conception du système de surveillance.

Un facteur de scénario a été choisi pour la nature du système cible (f_1) et trois facteurs de conception ont été choisis pour le SPOC et le processus expérimental (f_2 à f_4). Le facteur f_1 caractérise la vitesse de propagation de la maladie. Il s'agit du taux de transmission de la maladie entre deux unités épidémiologique voisines dans le réseau infectieux. Les facteurs f_2 et f_3 concernent des choix de conception du système d'épidémiosurveillance. Il s'agit respectivement de la période entre deux campagnes d'échantillonnage et de la précision relative désirée utilisée pour le calcul de la proportion d'unités épidémiologique à échantillonner lors de chaque campagne. Enfin, le facteur f_4 concerne la conception du processus expérimental. Il s'agit de déterminer si l'on doit disposer ou non de la valeur moyenne du taux de rétablissement d'une unité épidémiologique. Si l'on dispose de cette valeur (notons la tx_m), le processus expérimental l'utilise pour calculer le paramètre γ du modèle SIS ($\gamma = tx_m$) et seul le paramètre β du modèle SIS et son état initial sont calibrés sur la base des données de prévalence. Si l'on ne dispose pas de cette valeur, le paramètre γ est également calibré sur la base des données de prévalence.

Quatre critères ont été choisis pour évaluer le contrôle finalement effectué par le SPOC et le processus expérimental. Les deux premiers critères (c_1 et c_2) caractérisent la perte de production entraînée par la maladie et les deux derniers critères (c_3 et c_4) portent sur les coûts de la surveillance et du contrôle. Le premier critère (c_1) correspond à la durée de quarantaine durant laquelle la région est considérée comme infectée et est donc interdite d'exportation. Il se mesure comme étant l'inter-

1. Le modèle SIS est présenté dans l'état de l'art en tant qu'exemple de modèle de système dynamique (cf. section 2.3.2)

Fact.	Mesure	Signification	Type
f_1	Taux de transmission entre unités épidémiologiques	Vitesse de propagation de la maladie	Nature de A
f_2	Intervalle de temps entre deux campagnes d'échantillonnage du système d'épidémiosurveillance	Période d'échantillonnage	Conception de ρ_o
f_3	Précision relative désirée utilisée pour le calcul de la proportion d'unités épidémiologiques échantillonnées lors de chaque campagne	Précision relative désirée	Conception de ρ_o
f_4	Connaissance ou non de tx_m	Connaissance ou non de tx_m	Conception de ρ_m
Crit.			
c_1	Intervalle de temps entre la première détection de cas et la date à laquelle la surveillance proactive mesure une prévalence nulle	Durée de quarantaine	Perte de production
c_2	Intégrale de la prévalence pendant la durée de l'épidémie	Diminution de production des élevages	Perte de production
c_3	Nombre d'unités épidémiologiques observées pendant l'épidémie par le système de surveillance proactive	Effort de surveillance	Coût
c_4	Intégrale du pourcentage de restriction de mouvements sur la durée de l'épidémie	Effort de contrôle	Coût

TABLE 4.2: Critères et facteurs pour répondre aux questions sur T_{uc}

valle de temps entre la détection du premier cas et la date à laquelle la surveillance proactive mesure une prévalence nulle. Le second critère (c_2) représente la perte de production due à l'infection des élevages. Il se mesure comme l'intégrale de la prévalence réelle (par opposition à la prévalence estimée par la surveillance) de la maladie sur toute la durée de l'épidémie. Il représente donc la « proportion d'unités épidémiologique infectées jours » sur la totalité de l'épidémie. Le troisième critère (c_3) représente l'effort de surveillance. Il se mesure comme le nombre d'unités épidémiologiques observées pendant l'épidémie par le système de surveillance proactive. Enfin, le quatrième critère (c_4) représente l'effort de contrôle. Il se mesure comme l'intégrale du pourcentage de restrictions de mouvement sur la durée de l'épidémie.

Le tableau 4.2 synthétise l'ensemble des critères et des facteurs définis pour pouvoir répondre aux questions posées sur T_{uc} .

4.4 Modèle « *under control* » (T_{uc}^*)

L'intérêt de construire le modèle T_{uc}^* est double. Tout d'abord certains critères ou facteurs ne sont pas mesurables dans T_{uc} . C'est le cas par exemple pour les

Paramètre	valeur ^a
Graphe d'interactions	cf. figure 4.5 (b)
Nombre d'unités inf.	100
Taux d'infection (tx_{inf})	0.2 ut^{-1}
Période infectieuse (t_{inf})	15 ut

TABLE 4.3: Paramètres pour la simulation de la propagation de la maladie

^a. ut = unité temporelle.

critères c_2 ou c_4 qu'on ne peut connaître avec certitude. Ensuite, le modèle T_{uc}^* va nous permettre de réaliser un plan d'expériences dans lequel on va faire varier les valeurs des facteurs et observer l'influence de ces variations sur les critères. La figure 4.3 représente la structure du modèle T_{uc}^* .

4.4.1 Le modèle de système cible

Le système épidémiologique est décomposé d'un ensemble d'unités épidémiologiques reliées entre elles par un réseau de contacts infectieux. Chaque unité épidémiologique est modélisée par un modèle DEVS et le réseau de contacts infectieux est le graphe de connexion entre les ports de ces modèles DEVS. La figure 4.4 présente la structure du modèle de système épidémiologique.

Notons les caractéristiques principales de ce modèle. Tout d'abord au sujet du modèle d'unité épidémiologique. Il est composé de deux sous-modèles. Le modèle de l'infection interne d'une unité épidémiologique est un automate à états finis à deux états : l'état S (pour susceptible) et I (pour infectieux). Dans son état stable, l'automate est à l'état S. S'il est perturbé par un événement sur son port « infection », il passe à l'état I. Il va rester à l'état I pendant une certaine durée appelée « période infectieuse » (notée t_{inf}) puis retourner automatiquement à l'état S. L'état d'une unité épidémiologique va donc évoluer selon la séquence $S \rightarrow I \rightarrow S$. Le modèle de transmission est respectivement activé et désactivé via son port « état » suivant que le modèle d'infection interne entre dans l'état infectieux ou en sort. Lorsque le modèle de transmission est actif il envoie des messages d'infection sur ses ports « voisins » à des intervalles de temps tirés aléatoirement dans une distribution exponentielle négative dont le taux est un paramètre du modèle. On appelle ce taux le « taux d'infection » (noté tx_{inf}). Il correspond au facteur f_1 dont on veut tester l'influence sur le comportement de T_{uc}^* . Pour une compréhension plus intuitive on peut décrire ce comportement comme un processus à temps discret. Cela reviendrait à dire qu'à chaque pas de temps, un nœud voisin d'un nœud infectieux a une probabilité de la valeur du « taux d'infection » de devenir infectieux à son tour. La figure 4.5 et le tableau 4.3 présentent un exemple de paramétrage du modèle et les résultats de simulation associés.

Notons en outre qu'une position géographique (latitude, longitude) est donnée pour chaque unité infectieuse, et que les modèles d'unités infectieuses sont dotés de ports status?, status, et contrôle pour pouvoir interagir avec les systèmes de surveillance et de contrôle.

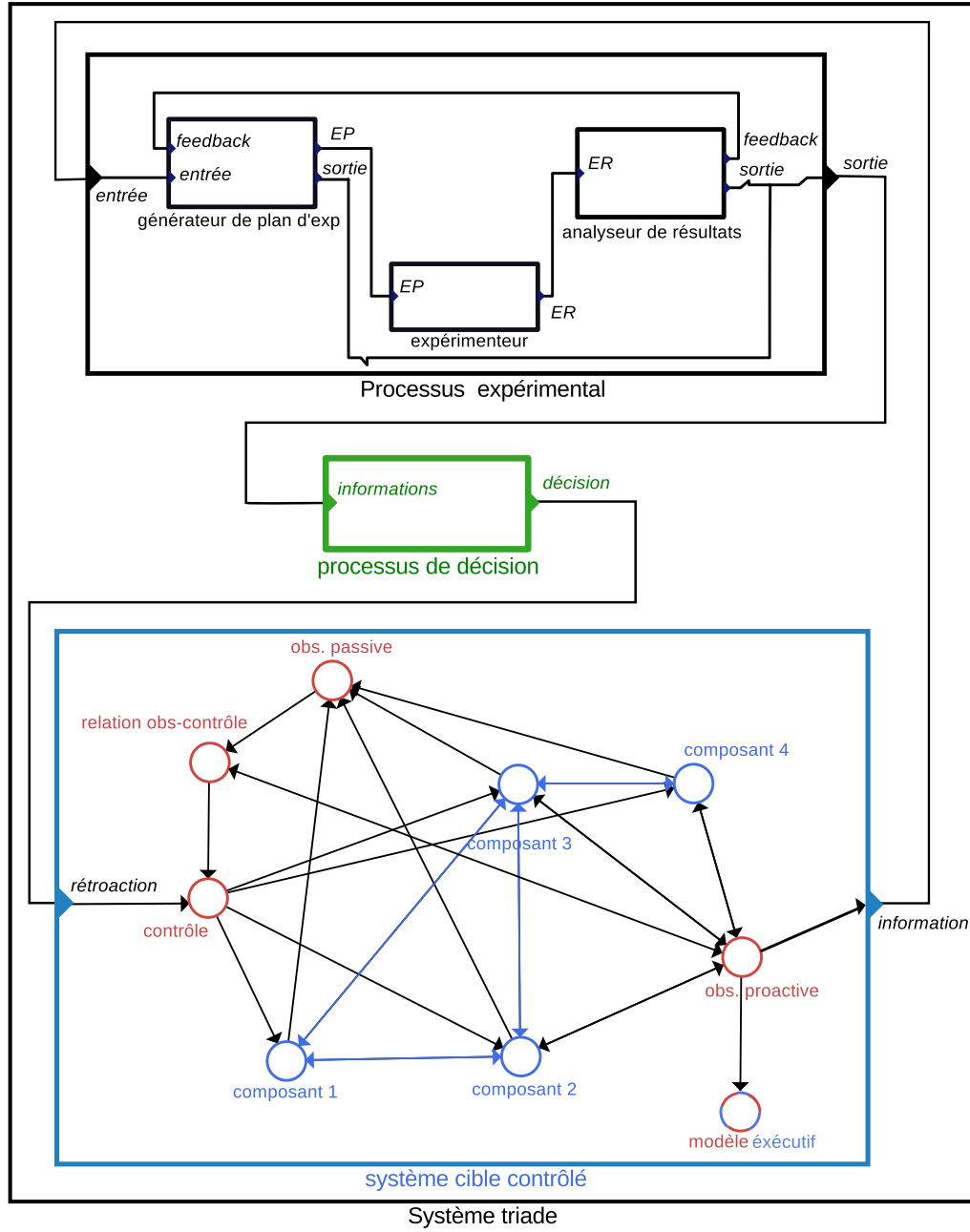


FIGURE 4.3: Structure du modèle T_{uc}^* . Les boîtes rectangulaires et circulaires représentent des modèles de systèmes dynamiques, spécifiés comme des modèles atomiques ou des modèles couplés. Les connexions sans flèches représentent des connexions « port → port ». Les connexions avec des flèches représentent une connexion « modèle → modèle » (c'est-à-dire une ou plusieurs connexions « port → port »). Les modèles représentant des processus du processus expérimental (en noir) se situent au niveau hiérarchique du modèle de processus expérimental. Les modèles représentant des processus du système cible (en bleu) et les modèles représentant des processus du SPOC (en rouge), se situent au niveau hiérarchique du système cible contrôlé. Le modèle de processus de décision (en vert) se situe au niveau hiérarchique de la triade. Le modèle exécutif du modèle de système cible contrôlé (en rouge et bleu) assure la gestion des connexions dynamiques entre les modèles du SPOC et les modèles composants du modèle de système cible.

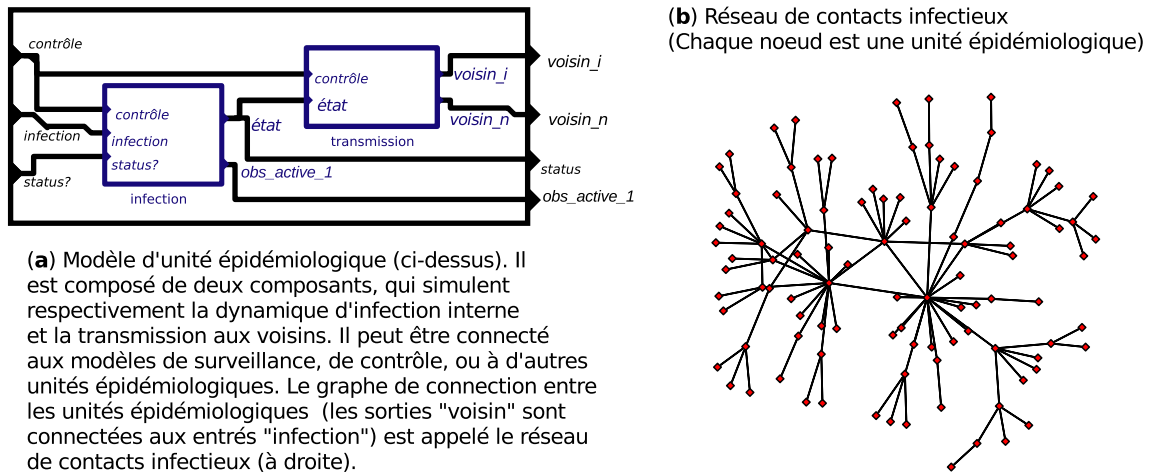


FIGURE 4.4: Modèle de propagation de la maladie : Il est lui même composé de modèles d'unités épidémiologiques (a) reliés ensemble par un réseau de contacts infectieux (b)

L'implémentation du modèle de système cible

Le modèle de système cible a été implémenté en utilisant la plateforme VLE. Le modèle d'unité épidémiologique a été défini comme une classe de modèles et l'ensemble du réseau de modèles est généré automatiquement à partir de cette classe et d'une matrice d'adjacence représentant la structure du réseau (Cf. annexe A.2). La structure du réseau de contact est donc facile à paramétrer.

Exemples d'utilisation du modèle de système épidémiologique

Deux applications du modèle de système épidémiologique indépendantes du modèle de système triade illustrent bien son fonctionnement [DBDG09, VDB⁺10]. Ces applications sont disponibles en annexe (cf. annexe I).

4.4.2 Le modèle de SPOC

Dans cette section, nous utilisons les modèles d'observation passive et d'observation proactive définis de manière générique dans le chapitre précédent (cf. chapitre 3 section 3.5.1). Ils sont couplés au modèle de système épidémiologique pour modéliser le système de surveillance épidémiologique décrit dans la présentation du système T_{uc} .

Modèle du système d'épidémiogigilance

Le système d'épidémiogigilance est modélisé par un modèle d'observation passive. Le modèle d'observation passive est connecté à tous les modèles d'unité épidémiologique. Les informations d'état qu'il reçoit sont à valeur dans $\{S, I\}$, soit l'unité épidémiologique est passée à l'état « S » soit à l'état « I » (cf. section 4.4.1). La fonction de traitement d'information nous sert à spécifier la probabilité qu'un cas soit détecté et que l'information soit transmise par le réseau d'épidémiogigilance si une

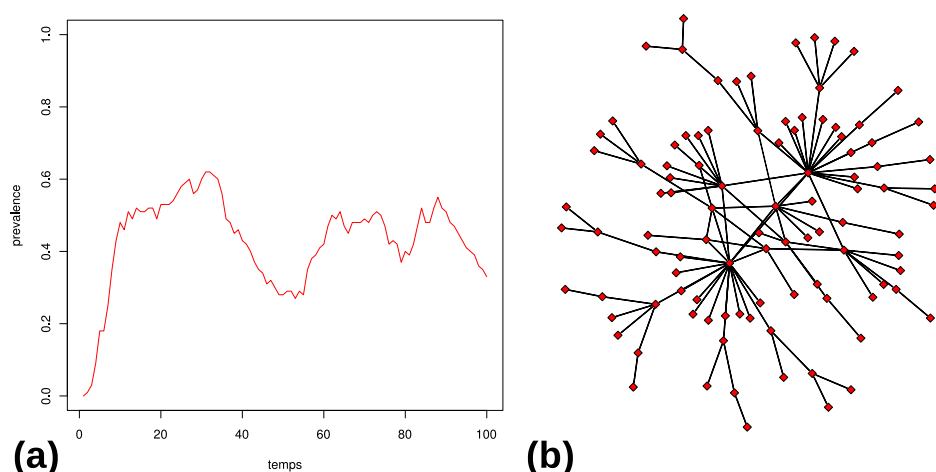


FIGURE 4.5: Exemple de simulation de propagation de la maladie. **(a)** Évolution de la proportion d'unités épidémiologiques susceptibles infectieuses au cours de la simulation. **(b)** Graphe d'interaction entre unités épidémiologiques pour cette simulation. Les paramètres utilisés sont présentés dans le tableau 4.3.

sous région est infectée. On note $p_{\text{declaration}}$ cette probabilité. Lorsqu'une information d'état « I » est reçue, le modèle d'observation passive transforme cette information en information « S » avec une probabilité de $1 - p_{\text{declaration}}$. Une fois l'information traitée, le test de transmission a lieu. Si l'état vaut « I » l'information est transmise. Sinon l'information n'est pas transmise.

Modèle du système d'épidémiosurveillance

Le système d'épidémiosurveillance est modélisé par un modèle d'observation proactive. Le modèle d'observation proactive est paramétré par une période d'observation notée « t_{obs} » (elle correspond au facteur f_2 du tableau 4.2), par la liste des identifiants des modèles d'unité épidémiologique, et par la précision relative désirée notée « ϵ » (elle correspond au facteur f_3 du tableau 4.2). La fonction de choix de modèles à échantillonner réalise un tirage aléatoire dans la liste d'identifiants des modèles composants. La fonction de traitement de l'information ne modifie pas l'information et la fonction d'agrégation de l'information calcule la prévalence observée dans l'échantillon. C'est-à-dire qu'elle consiste à diviser le nombre d'états « I » reçus par la taille de l'échantillon. La fonction de transmission consiste à toujours transmettre la prévalence calculée. Notons que la fonction de choix de modèles à échantillonner doit également spécifier la taille de l'échantillon (N). La valeur de N est calculée à chaque pas de temps à l'aide des équations 4.1 et 4.2. Ces formules, basées sur des lois statistiques, sont utilisées en science vétérinaires pour calculer des tailles d'échantillon à partir de la précision relative désirée et de la prévalence supposée [DMS10]².

2. Il s'agit des équations 2.4 et 2.8 de l'ouvrage [DMS10].

$$N_{\infty} = 1.96^2 \frac{p_0(1-p_0)}{(\epsilon.p_0)^2} \quad (4.1)$$

$$N = \frac{1}{1/N_{\infty} + 1/N_{pop}} \quad (4.2)$$

Où

p_0 est la prévalence observée au pas de temps précédent. (À l'initialisation, on pose $p_0 = 0.01$.),

N_{∞} est la taille que devrait avoir l'échantillon pour une population infinie,

N est la taille de l'échantillon pour une population de taille N_{pop} ,

ϵ est l'erreur relative acceptée.

Une nouvelle fonctionnalité a été apportée au modèle d'observation proactive par rapport à la description qui en est faite dans le chapitre 3. Deux ports d'entrée ont été ajoutés, un port « start » et un port « stop ». La réception d'un évènement sur le port « start » active le modèle et la réception d'un évènement le port « stop » le désactive. Cela permet d'activer le modèle d'observation proactive lorsque le premier cas est détecté et de le désactiver lorsque l'épidémie est terminée.

Modèle de contrôle

Une restriction des mouvements d'animaux de niveau « i » est modélisée par une réduction de $x_i\%$ du taux d'infection des unités épidémiologiques. Un évènement de restriction des mouvements contient un attribut booléen « on/off » disant si la mesure de contrôle est mise en place ou supprimée, et une intensité $x\%$ qui correspond au ratio de réduction. Si une unité épidémiologique reçoit un évènement de restriction des mouvements portant les attributs $x_i\%$ et « on/off = 1 » sur son port « contrôle », le modèle de transmission diminue son attribut « taux d'infection » de $x_i\%$ (i.e. si on note tx_{inf} le taux d'infection, $tx_{inf} \leftarrow tx_{inf} - x_i.tx_{inf}$). Si « on/off = 0 », il rétablit l'ancien taux d'infection ($tx_{inf} \leftarrow tx_{inf}/(1 - x_i)$).

L'envoi des évènements de restriction de mouvement est assuré par le modèle de contrôle. Le modèle de contrôle possède un port d'entrée (« décision ») et un port de sortie (« contrôle »). Son port de sortie est connecté à tous les modèles composants sur leur port d'entrée « contrôle » (cf.figure 4.4). Son port d'entrée est connecté au modèle de processus de décision et au modèle de relation observation contrôle (cf.figure 4.6). Les évènements reçus par le modèle de contrôle doivent contenir une intensité de contrôle (un pourcentage) et un attribut « on/off ». Lorsque le modèle de contrôle reçoit un évènement, il patiente pendant un certain délai (que l'on notera d_c) puis transmet l'évènement de restriction de mouvements aux unités épidémiologiques.

Modèle couplé de SPOC

La figure 4.6 présente le schéma des connexions entre les modèles composant le modèle de SPOC. Le modèle nommé « relation controle observation » a du être ajouté pour que la chronologie des actions de contrôle et d'observation réalisées par

SPOC correspond bien à la chronologie de la situation d'épidémie contrôlée présentée figure 4.2. Lorsque le modèle de relation contrôle observation reçoit un message sur son port « start », il envoie un message contenant la valeur par défaut du ratio de restriction de mouvements ($x_0\%$) sur son port « control ». Cela entraîne l'activation du modèle de surveillance proactive et du modèle de contrôle. Il reçoit ensuite toutes les valeurs de prévalence mesurées par le modèle d'observation proactive sur son port « continue ». Si il reçoit plus de 5 valeurs de prévalence nulles successives, il envoie un évènement sur son port « stop ». Cette action désactive le modèle d'observation proactive.

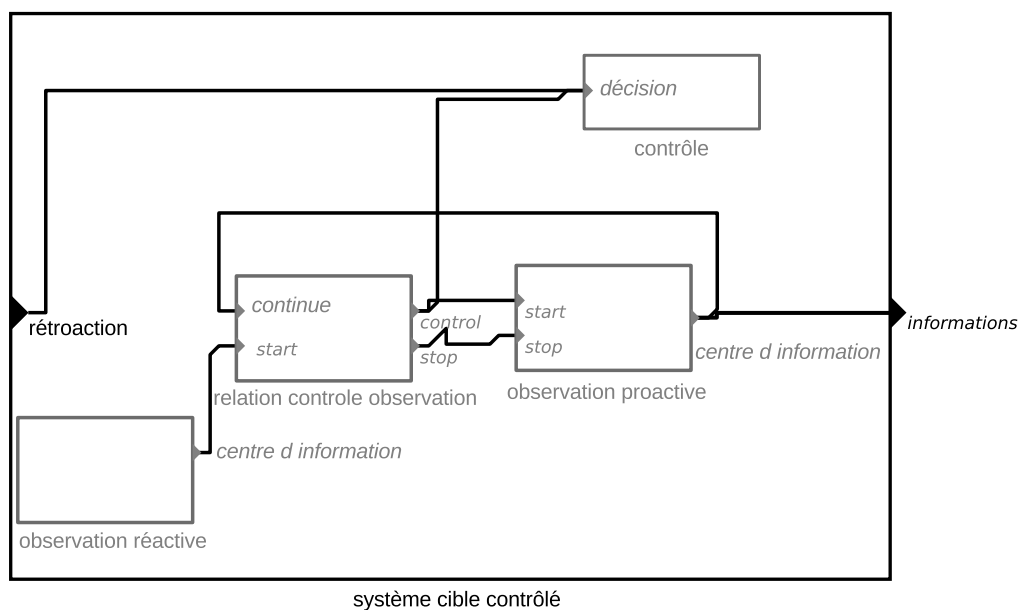


FIGURE 4.6: Schéma des connexions entre les modèles composant le modèle de SPOC de T_{uc}^*

Modèle de système cible contrôlé

La structure du modèle de système cible contrôlé est illustrée dans la figure 4.3 page 84. Notons que le modèle de système cible contrôlé est le modèle d'une épidémie contrôlée, surveillée par un système de surveillance épidémiologique. Ce modèle peut être utilisé en tant que modèle A* d'une triade de Minsky dont l'objet A serait un système épidémiologique surveillé et contrôlé [BD11]. Une description de ce type d'utilisation est disponible en annexe pour le lecteur intéressé (cf. annexe J).

4.4.3 Le modèle de processus expérimental

Le processus expérimental du système T_{uc} a pour but de trouver la valeur de R_0 à associer au système épidémiologique contrôlé. Nous le modélisons ici par un modèle MPE_{OEP} tel qu'il est décrit dans l'annexe D. Nous l'avons paramétré selon les indications données dans [Ped10].

Le modèle A^{} cadré**

Dans cet exemple, le modèle A^{**} que nous avons choisi pour représenter le modèle A_B^* est en fait le même objet que A_B^* : un modèle SIS. Rappelons que notre motivation première de modéliser la triade de Minsy est la non répétabilité d'une expérience sur le système cible. En construisant un modèle de la triade, nous nous donnons la possibilité d'expérimenter le modèle triade. Ce qui nous intéresse est d'évaluer la pertinence d'utilisation du modèle A_B^* dans un cadre que l'on puisse expérimenter. Notons toutefois que le modèle A^{**} n'est pas systématiquement égal au modèle A_B^* pour toutes les triades T . Dans un cas de figure où l'on voudrait trouver le meilleur modèle par exemple, il faudrait tester de nombreux objets A^{**} qui peuvent être par exemples des simplifications du modèle A_B^* .

Le modèle SIS est un MSD couramment utilisé en épidémiologie : le modèle SIS (cf. équations 4.3, et 4.4). Ce modèle est présenté en tant qu'exemple de modèle dynamique déterministe dans la section 2.3.2 du chapitre 2.

$$\frac{dS}{dt} = \gamma I - \beta IS \quad (4.3)$$

$$\frac{dI}{dt} = \beta IS - \gamma I \quad (4.4)$$

Où

$\{S, I\} \in \mathbb{R}^2$ sont deux variables d'états (respectivement le nombre d'individus susceptibles et infectieux),

β et γ sont deux paramètres (respectivement le taux d'infection et le taux de rétablissement).

Un intégrateur de type QSS a été utilisé pour formaliser le modèle SIS en DEVS (cf. section 2.7.4 page 45). Une implémentation d'intégrateur QSS générique est fournie dans la plateforme VLE qui impose une spécification particulière des modèles de type QSS. Ces modèles possèdent deux ports d'entrée « perturb » et « request » qui permettent respectivement de modifier les valeurs des variables d'états ou de certains paramètres du QSS et de demander les valeurs des variables d'états à un instant donné. Si une demande d'information sur une variable d'état arrive sur le port « request », la valeur à cette date est calculée (la demande peut avoir lieu entre deux pas d'intégration) et renvoyée sur le port « response ». De plus le modèle QSS possède un port de sortie pour chacune de ses variables d'états (du nom de la variable) sur lesquels il dépose les valeurs respectives de ses variables à chaque pas d'intégration. Un intégrateur QSS pour un système à deux variables d'états (S et I dans notre cas) doit être paramétré par trois paramètres, une valeur de seuil (s) en dessous de laquelle un taux de variation est considéré comme nul, et deux paramètres de pas d'intégration ($steps_S$ et $step_I$) donnant les pas d'intégration pour chacune des dimensions dans l'espace d'état. Le modèle SIS devient le modèle A^{**} de la triade T_{uc}^* . Il s'accommode à un cadre expérimental CE_C tel qu'il est décrit dans l'annexe D. La structure du modèle couplé composé d'un cadre expérimental de type CE_C et du modèle SIS est présentée figure 4.7. Dans notre exemple nous testons en fait deux cadres expérimentaux de calibration différents : CE_{C1} et CE_{C2} . Le premier perturbe uniquement la valeur de I et de β à l'initialisation. Le second

perturbe la valeur de I , de β et de γ . De cette manière il est possible de tester l'influence du facteur f_4 (cf. tableau 4.2). Le cadre CE_{C1} correspond au cas où γ est connu, auquel cas la bonne valeur est utilisée par défaut dans le modèle SIS et le MPE_{OEP} doit répondre à un problème en deux dimensions (il faut trouver β et I_{init}). Le cadre CE_{C2} correspond au cas contraire où γ est inconnu, auquel cas le MPE_{OEP} doit répondre à un problème en trois dimensions (il faut trouver β , I_{init} et γ).

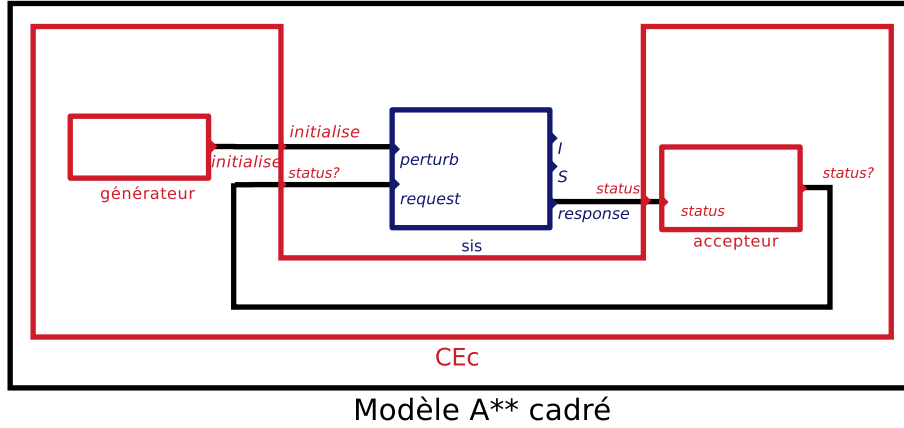


FIGURE 4.7: Cadre expérimental et modèle utilisés par le MPE de T_{uc}^*

Ajoutons une précision sur l'utilisation du paramètre β . Le coefficient qui multiplie la variable S dans le terme de transmission du modèle SIS ($\pm\beta IS$) est appelé la force d'infection ($\lambda = \beta I$ dans notre cas). La signification originale du paramètre β dans cette force d'infection est expliquée en fonction du taux de contact entre les individus de la population et de la probabilité d'infection lors d'un contact entre un individu infectieux et un individu susceptible (cf. [KR07, Box 2.1]). Cependant, cette explication suppose notamment un taux de contact égal entre tous les individus de la population, ce qui n'est pas respecté par certaines configurations du réseau d'infection du système cible. Pour cette raison, dans l'utilisation que nous faisons du modèle SIS, le paramètre β est compris comme étant un coefficient d'accroissement linéaire de la force d'infection en fonction du nombre d'individus infectieux dans l'approximation la plus simple possible qui consiste à modéliser la force d'infection comme étant proportionnelle au nombre d'individus infectieux.

Le MPE_{OEP} (ρ_m^*)

Le MPE_{OEP} (cf. annexe D) du modèle T_{uc}^* est connecté au modèle de système cible contrôlé et au modèle de processus de décision selon le schéma de connexion présenté figure 4.3 page 84. La fonction de déclenchement utilisée consiste à déclencher une calibration dès lors qu'un nombre minimal de données ont été reçues (Nb_{obs}). La fonction de distance utilisée est l'écart moyen (cf. fonction Mo dans le tableau 2.1). Enfin le critère d'arrêt est un nombre d'itérations maximal noté Nb_{it} . L'espace de recherche (en deux ou trois dimensions selon la valeur de f_1) est redéfini à chaque calibration. Les bornes inférieures et supérieures de la valeur de I_{init} (la valeur de la variable I à l'initialisation du modèle SIS) sont recalculées en fonction de

la dernière estimation de prévalence reçue et de l'erreur relative désirée. Si l'on note p_0 la dernière valeur de prévalence reçue, les bornes inférieures et supérieures (notées respectivement I_{inf} et I_{sup}) sont calculées de la manière suivante : $I_{inf} = p_0 - \epsilon.p_0$ et $I_{sup} = p_0 + \epsilon.p_0$. Après une phase de calibration, la série temporelle des observations est réinitialisée.

4.4.4 Modèle de processus de décision

Le processus de décision consiste à augmenter ou diminuer l'intensité de la mesure de restriction des mouvements d'animaux en fonction de la valeur de R_0 estimée par le processus expérimental. Le modèle de processus de décision comprend un port d'entrée « informations » sur lequel il reçoit les valeurs de paramètres β et γ estimées par le modèle de processus expérimental, et un port de sortie « décision » sur lequel il envoie la nouvelle intensité de restriction des mouvements (un pourcentage). Il est paramétré par un vecteur de pourcentages $(x_1, \dots, x_n) \in [0, 1]^n$ qui correspondent aux pourcentages de diminution du taux d'infection des unités épidémiologiques pour chacun des « n » niveaux d'intensité de la restriction de mouvements. Décrivons le comportement du modèle de processus de décision à la réception d'un événement d'information.

Imaginons que le niveau d'intensité précédent était le niveau $j \in [1, n - 1]$. La valeur de R_0 est calculée par la formule suivante : $R_0 = \beta/\gamma$. Trois cas de figure sont alors possible :

- si $R_0 > 1$, le niveau d'intensité supérieur est enclenché, deux événements sont envoyés au modèle de contrôle. Le premier contient un ratio de x_j et un attribut « on/off = 0 ». Le second contient un ratio de x_{j+1} et un attribut « on/off = 1 » ;
- si $R_{0_tol} < R_0 \leq 1$ aucun événement n'est envoyé ;
- si $R_0 < R_{0_tol}$, on redescend au niveau d'intensité inférieur, deux événements sont envoyés au modèle de contrôle. Le premier contient un ratio de x_j et un attribut « on/off = 0 ». Le second contient un ratio de x_{j-1} et un attribut « on/off = 1 ».

R_{0_tol} correspond à une valeur de R_0 qui serait tolérée par un décideur. Dans notre modèle nous l'avons fixée à 0.01 qui est la plus petite valeur de R_0 calculable par le modèle de processus expérimental. $R_0 = 1$ correspond à la valeur théorique en dessous de laquelle l'espérance du nombre d'individus infectés à l'équilibre est nul pour un système SIS (cf. chapitre 2, section 2.5.1), c'est-à-dire que la maladie ne peut pas devenir endémique.

4.5 Expérimentation de T_{uc}^*

L'objectif des simulations (ou expériences in-silico) du modèle T_{uc}^* est de tester l'influence des facteurs f_1 à f_4 sur les critères c_1 à c_4 dans un système représentatif du système T_{uc} . Nous avons vu dans les sections précédentes (sections 4.4.1 à 4.4.3) comment le modèle T_{uc}^* a été construit pour représenter le système T_{uc} . Dans cette section nous présentons les résultats d'un plan d'expérience visant à mesurer l'influence des facteurs f_1 à f_4 sur les critères c_1 à c_4 pour une paramétrisation donnée du modèle T_{uc}^* .

4.5.1 Une expérience du modèle T_{uc}^*

Les paramètres utilisés pour les expériences du modèle T_{uc}^* sont présentés dans les tableaux 4.4 et 4.5. Le modèle T_{uc}^* est stochastique et donc il est possible de générer des expériences différentes avec un même jeu de paramètres en changeant la graine du générateur de nombres aléatoires. Observons les résultats d'une expérience choisie dans le plan d'expériences. La figure 4.8 présente les sorties de simulation observées pour une expérience pour laquelle les facteurs prennent les valeurs suivantes : $f_1 = 0.045 \text{ ut}^{-1}$, $f_2 = 5 \text{ ut}$, $f_3 = 40\%$ et $f_4 = 0$. La simulation dure 200 unités de temps.

Trajectoire du modèle de système cible

Sur le graphique a) on peut voir l'évolution de la prévalence dans le modèle de système épidémiologique au cours du temps (courbe pleine). La courbe pointillée montre un exemple d'évolution de la prévalence pour une simulation du même modèle de système épidémiologique avec les mêmes valeurs de paramètres mais totalement déconnecté du modèle de SPOC. On observe que dans le cas où le système est contrôlé par le SPOC l'épidémie est maîtrisée (la prévalence est nulle en fin de simulation pour la courbe pleine), alors que dans le cas où elle n'est pas contrôlée, la maladie devient endémique (l'évolution de la prévalence semble se stabiliser autour d'une valeur non nulle pour la courbe pointillée).

Trajectoire perçue par le modèle de SPOC

Le graphique b) montre les valeurs de prévalence estimées par le modèle d'observation proactive. Il s'agit de la prévalence observée dans les échantillons (points noirs). La prévalence « réelle » dans le système épidémiologique a été superposée (courbe rouge) et on peut observer que l'estimation est très proche. L'espacement dans le temps entre les points de prévalence estimée correspond à la période d'échantillonnage du système d'épidémiosurveillance. La date de la première estimation de prévalence correspond à la date d'activation du modèle d'observation proactive déclenchée par la première détection du modèle d'observation passive. On remarque qu'elle a lieu un peu après le début de l'épidémie (la courbe rouge a déjà commencé à croître).

Trajectoire d'un indicateur du fonctionnement du système d'observation

Le graphique c) représente la proportion d'unités épidémiologiques échantillonnée à chaque échantillonnage réalisé par le modèle d'observation proactive (rappelons que le nombre d'unité à échantillonner est calculé par le modèle d'observation proactive d'après les équations 4.1 et 4.2), on peut voir ici que les variations sont importantes.

Informations apportées par le processus d'expérimentation de modèle modélisé dans le modèle T_{uc}^*

Le graphique d) illustre les résultats obtenus par le modèle de processus expérimental. Les données de prévalence estimées par la surveillance ont été tracées à

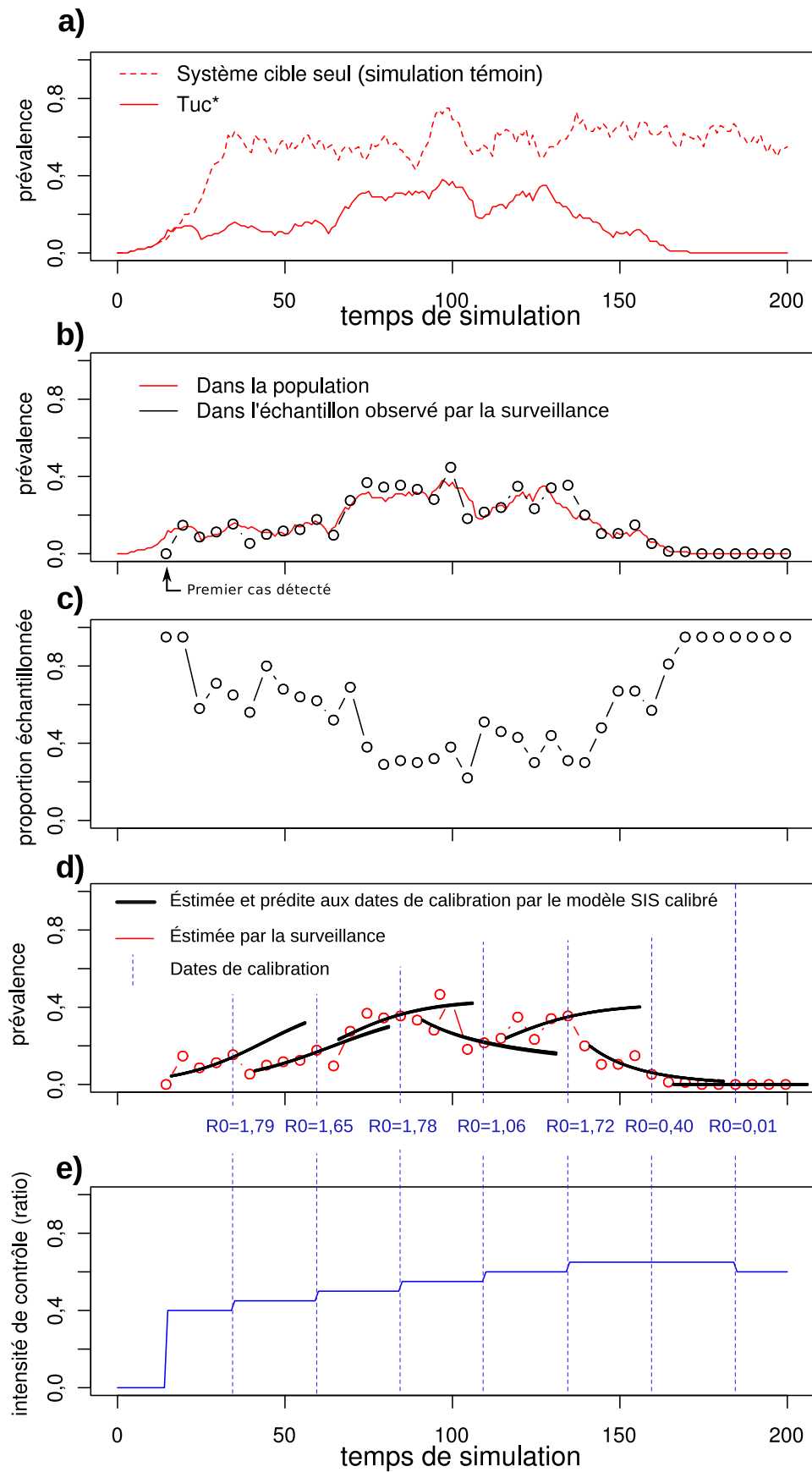


FIGURE 4.8: Sorties de simulations du modèle T_{uc}^* (simulation pour $f_1 = 0.045 \text{ ut}^{-1}$, $f_2 = 5 \text{ ut}$, $f_3 = 40\%$ et $f_4 = 0$). Cf. explication section 4.5.1.

nouveau (points rouges). Rappelons que les calibrations se font sur la base de ces données. Les barres pointillées bleues verticales marquent les dates auxquelles ont eu lieu les calibrations (i.e. à chaque fois que 5 données de prévalence étaient disponibles). Pour chaque calibration, l'évolution de la prévalence simulée par le modèle SIS calibré a été tracée depuis la date de la première observation utilisée pour cette calibration (5 observations avant) jusqu'à un horizon de prédiction fixé arbitrairement de manière à être inférieur à la durée entre deux calibrations (il s'agit des segments de courbes noires). La valeur de R_0 estimée lors de chaque calibration est inscrite en bleu en dessous de chaque date de calibration. On remarque que les prévalences simulées avec le modèle SIS s'accordent très bien aux données (les 5 observations de prévalence précédant une calibration sont presque superposées à la courbe noire correspondante), mais prédisent mal la prévalence à venir (les observations suivant une phase de calibration peuvent être très éloignées de la courbe noire correspondant à cette calibration). Ces observations ne sont pas surprenantes puisque le modèle SIS calibré à une date de calibration donnée est calibré sur une épidémie contrôlée à un certain niveau de contrôle et qu'à partir de la date de calibration, le niveau de contrôle change.

Trajectoire d'un indicateur du fonctionnement du système de contrôle

Le graphique e) présente l'évolution de l'intensité de contrôle (valeur de x_i) au cours de la simulation. On remarque tout d'abord que le modèle de contrôle est activé en même temps que le modèle de surveillance proactive (intensité de contrôle à 0 jusqu'à la première valeur de prévalence observée sur le graphique d). À chaque phase de calibration le niveau de contrôle est réévalué. Il est augmenté si le R_0 estimé est supérieur à 1 (ce qui est le cas jusqu'à l'antépénultième calibration), maintenu à sa valeur précédente si $R_{o_tol} < R_0 < 1$ (ce qui est le cas à l'avant dernière calibration), et diminué si $R_0 \leq R_{o_tol}$ ce qui est le cas de la dernière calibration.

Valeurs des paramètres et calcul des critères

On peut voir dans le tableau 4.4 que les facteurs font partie des paramètres du modèle. Les critères quant à eux peuvent être calculés sur la base d'une expérience. Les définitions de la mesure de chaque critère (c_1 à c_4) données dans le tableau 4.2 sont directement calculables sur la base des sorties de simulations que nous venons de présenter. La figure 4.9 illustre de manière graphique la mesure de chacun des critères pour l'expérience présentée dans cette section.

4.5.2 Plan d'expérience

Le plan d'expérience, a consisté à réaliser 30 simulations pour chaque combinaison possible des modalités choisies pour chacun des facteurs (on parle d'un plan d'expérience exhaustif [STCR04]). Ces modalités sont présentées dans le tableau 4.5. Dans le but de simplifier les discussions, seules deux modalités pour les facteurs f_1 et f_2 ont été retenues. Trois modalités ont été testées pour le facteur f_3 pour montrer l'intérêt d'un nombre de modalités supérieur à 2. Les valeurs de paramètres utilisés sont celles présentées dans le tableau 4.4. L'ensemble des simulations (720 simulations) a été réalisé en une vingtaine d'heures de calcul sur un ordinateur de bureau

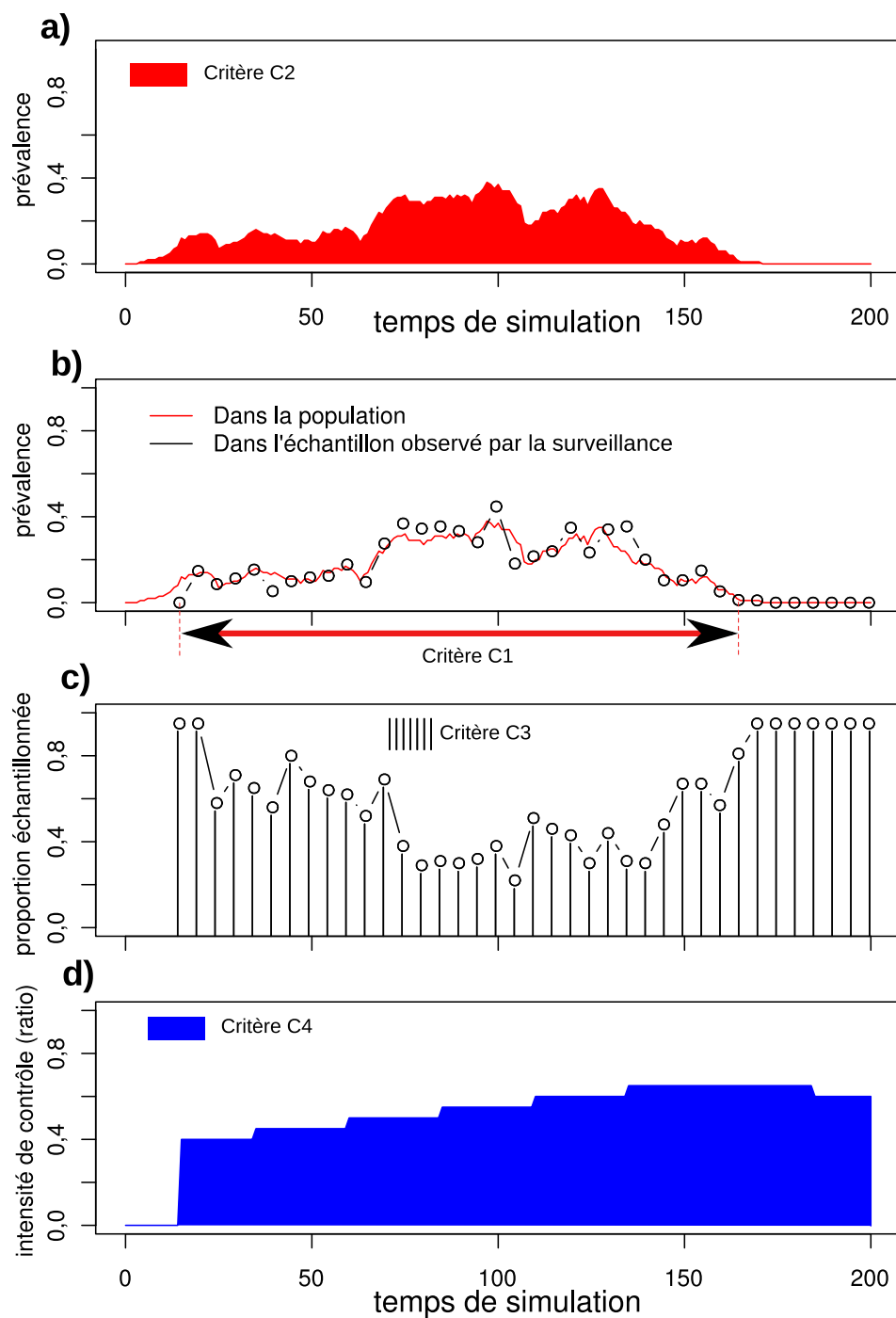


FIGURE 4.9: Illustration du calcul des critères c_1 à c_4 à partir des sorties de simulation de T_{uc}^* (cf. section 4.5.1).

Modèle	Paramètre	Signification	Valeur ^a
Système épidémio. (A_C^*)	tx_{inf}	Taux d'infection	f_1^b
	t_{inf}	Période infectieuse	10 ut
	g	Réseau infectieux	Lattice (voisinage de Moore) ^c
	N_{inf}	Nombre initial d'unités infectées	1
	N_{sus}	Nombre initial d'unités susceptibles	99
SPOC (ρ_o^*)	t_{obs}	Période d'observation	f_2
	ϵ	Précision relative désirée	f_3
	(x_1, \dots, x_n)	Intensités des niveaux de contrôle	0.4 à 0.7 par pas de 0.05
	$p_{declaration}$	Probabilité de déclaration	0.05
	d_c	Délai de contrôle	0.5 ut
Décision (B^*)	R_{0_max}	Seuil de R_0	1
	R_{0_tol}	R_0 toléré	0.01
EPM _{OEP} (ρ_m^*)	ω	Paramètre OEP	0.8 si $f_4 = 1$; 0.3925 si $f_4 = 0$
	ϕ_g	Paramètre OEP	1.3358
	ϕ_p	Paramètre OEP	2.5586
	N	Nombre de particules OEP	25
	N_{it}	Nombre d'itérations OEP	100 si $f_4 = 1$; 200 si $f_4 = 0$
	b_{low}	Borne inf. de l'espace de recherche OEP	$(I_{inf}, 0.001)$ si $f_4 = 1$; $(I_{inf}, 0.001, 0.001)$ si $f_4 = 0$ ^d
	b_{up}	Borne sup. de l'espace de recherche OEP	$(I_{sup}, 0.9999)$ si $f_4 = 1$; $(I_{sup}, 0.9999, 0.9999)$ si $f_4 = 0$
	CE_C	Cadre expérimental de calibration	CE_{C1} si $f_4 = 1$; CE_{C2} si $f_4 = 0$
	M_C	Modèle à calibrer (A^{**})	SIS
	Nb_{obs}	Nombre d'observations minimal	5
SIS (A^{**})	$steps_S$	Pas d'intégration selon S	10^{-4}
	$steps_I$	Pas d'intégration selon I	10^{-4}
	s	Seuil	10^{-13}
	β	Taux d'infection	calibré
	I_{init}	Proportion infectés initiale	calibré
	γ	Taux de rétablissement	$1/t_{inf}$ si $f_4 = 1$; calibré si $f_4 = 0$

TABLE 4.4: Paramètres de T_{uc}^*

a. ut = unité temporelle.

b. Les valeurs choisies pour les facteurs f_1 à f_4 sont données dans le tableau 4.5.

c. Une lattice traduit la relation de contact observée entre les cases d'une grille. Pour un voisinage de Moore chaque case est en contact avec 8 voisins (à gauche, à droite, en haut, en bas et sur les 4 diagonales).

d. Cf. section 4.4.3 pour le calcul de I_{inf} et I_{sup}

Facteur	Modalité	Valeur ^a
Vitesse de propagation de la maladie (f_1)	Lente	0.04
	Rapide	0.045
Période d'échantillonnage (f_2)	Courte	3 ut
	Moyenne	4 ut
	Longue	5 ut
Précision relative désirée (f_3)	Précise	20%
	Peu précise	40%
Connaissance de γ (f_4)	Oui	1
	Non	0

TABLE 4.5: Plan d'expériences

^a. ut = unité temporelle.

64 bits à 4 processeurs (processeurs Intel 2.66 GHz) avec 4 Go de mémoire RAM. Le lancement des simulations pouvant être automatisé et le logiciel VLE pouvant être installé sur un cluster de calcul, il est possible de réaliser des plans d'expérience exhaustifs contenant davantage de simulations et de tester notamment différentes valeurs de paramètres ou davantage de modalités pour les facteurs correspondant à des variables continues (facteurs f_1 à f_3).

4.5.3 Analyse des résultats d'expérience

L'analyse des résultats de simulations est orientée par les questions posées au système T_{uc} dans la section 4.2.2. Nous transférons ici ces questions au modèle T_{uc}^* . La figure 4.10 illustre les relations entre les critères pour l'ensemble des expériences.

Impact et amélioration de la surveillance

Le tableau 4.6 présente les résultats d'une analyse de la variance du critère c_2 (perte de production élevage) en fonction de chacun des facteurs. On peut voir que choisir une modalité plutôt qu'une autre pour la période d'échantillonnage (facteur f_2) a un rôle significatif sur le critère c_2 . Au contraire, la modification de la précision d'échantillonnage désirée (facteur f_3) que nous avons testé (passer d'une précision relative de 20% à 40%) n'a pas d'influence significative. Les mêmes résultats qualitatifs sont observés pour le critère c_1 qui représente la durée de quarantaine (cf. tableau H.1 annexe H). Le tableau 4.7 présente les coefficients du modèle linéaire utilisé pour l'analyse de la variance. On peut voir qu'une augmentation de la période d'échantillonnage (à $f_2 = 4$ ut ou $f_2 = 5$ ut par rapport à un intercept à $f_2 = 3$ ut) entraîne en moyenne une augmentation du critère c_2 puisque les coefficients estimés sont positifs. En fait une augmentation de f_2 de 3 ut à 4 ut n'entraîne pas d'augmentation significative de la valeur de c_2 (0.7899 %.ut) alors qu'une augmentation de 3 ut à 5 ut de f_2 entraîne une augmentation significative de c_2 (4.8846%.ut) qui est plus de 6 fois supérieure (0.7899/4.8846) que l'augmentation entraînée par le passage de 3 ut à 4 ut.

Les critères c_1 et c_2 visent à estimer la perte de production entraînée par la maladie. Nous venons de voir que, pour les modalités testées, le facteur f_2 (période

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Test ^a
f1	1	16009	16008.7	141.5921	< 2.2e-16	***
f2	2	3300	1650.0	14.5939	6.139e-07	***
f3	1	14	13.9	0.1229	0.7259954	
f4	1	1315	1315.3	11.6330	0.0006844	***
Residus	714	80726	113.1			

TABLE 4.6: Analyse de la variance du critère c_2 (diminution de production dans les élevages) en fonction des critères f_1 à f_4 .

a. Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

d'échantillonnage) a une influence significative sur ces critères tandis que le facteur f_3 (précision relative) n'en a aucune. Observons maintenant les influences des facteurs f_1 et f_2 sur les critères permettant d'estimer le coût des mesures de surveillance et de contrôle. Il s'agit des critères c_3 et c_4 qui mesurent respectivement le nombre d'unités épidémiologiques observées au cours de l'épidémie et l'intégrale du ratio de contrôle au cours de l'épidémie. On peut observer (dans le tableau 4.8) que le facteur f_3 , qui n'a pas d'influence significative sur les critères de perte de production, a une influence significative sur l'effort de surveillance (critère c_3), et n'a pas d'influence significative sur l'effort de contrôle (critère c_4 tableau 4.9). Réduire la précision désirée de l'échantillonnage de 20% à 40% permet donc pour les modalités testées de diminuer significativement le critère estimant l'effort de surveillance (c_3), sans augmenter de manière significative les critères d'estimation des coûts de contrôle (c_4) ou de perte de production due à la maladie (c_1 et c_2).

Coût de la surveillance Le critère c_3 , qui nous sert à estimer le coût de la surveillance, pourrait au premier abord sembler découler directement des facteurs f_2 (période d'échantillonnage) et f_3 (précision relative) qui concernent la conception du système de surveillance. L'impact de la valeur des facteurs f_2 et f_3 sur le critère c_3 est en fait très complexe. Une diminution de la période d'échantillonnage (f_2) par exemple, tendrait à faire augmenter le critère c_3 puisque davantage d'échantillons sont récoltés pendant le même laps de temps. Cependant, on a montré qu'une diminution de f_2 diminuait de manière significative la durée de la surveillance (critère c_2), ce qui tendrait à diminuer la valeur du critère c_3 . On peut observer dans les résultats d'expérience que l'augmentation du critère c_3 n'est pas proportionnelle à l'augmentation du facteur f_2 en regardant les valeurs de coefficients du modèle linéaire $c_3 \sim f_1 + f_2 + f_3 + f_4$ estimées pour les différentes modalités de f_2 (cf. annexe H, tableau H.3).

Impact du cadre expérimental utilisé pour le modèle SIS

On peut voir dans le tableau 4.6 que le fait de disposer ou non du paramètre γ du modèle SIS (facteur f_4) a un impact significatif sur le critère c_2 (perte de production élevage). Le critère c_2 est augmenté si γ n'est pas connu. Cependant on peut voir également que l'augmentation du critère c_2 due à l'ignorance de γ est deux fois plus faible que l'augmentation due à un passage d'une période d'échantillonnage de 3 ut à 5 ut (comparaison des lignes « f4 0 » et « f2 5 » du tableau 4.7). Les autres critères

	Estimate	Std.Error	t value	Pr(> t)	Test ^a
(Intercept)	11.3800	0.9707	11.724	<2e-16	***
f1 0.045	9.4306	0.7925	11.899	<2e-16	***
f2 4	0.7899	0.9707	0.814	0.416035	
f2 5	4.8846	0.9707	5.032	6.14e-07	***
f3 0.4	-0.2779	0.7925	-0.351	0.725995	
f4 0	2.7031	0.7925	3.411	0.000684	***

TABLE 4.7: Coefficients modèle linéaire $c_2 \sim f_1 + f_2 + f_3 + f_4$

a. Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Test ^a
f1	1	12161	12161	33.1449	1.270e-08	***
f2	2	47810	23905	65.1558	< 2.2e-16	***
f3	1	36982	36982	100.7984	< 2.2e-16	***
f4	1	200	200	0.5456	0.4604	
Residuals	714	261962	367			

TABLE 4.8: Analyse de la variance du critère c_3 (coût de surveillance) en fonction des critères f_1 à f_4 .

a. Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Test ^a
f1	1	65092	65092	86.8170	<2.2e-16	***
f2	2	8316	4158	5.5457	0.004074	**
f3	1	5	5	0.0065	0.935695	
f4	1	60	60	0.0798	0.777706	
Residuals	714	535331	750			

TABLE 4.9: Analyse de la variance du critère c_4 (coût de contrôle) en fonction des critères f_1 à f_4 .

a. Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Estimate	Std.Error	t value	Pr(> t)	Test ^a
(Intercept)	14.0396	1.2434	11.292	<2e-16	***
f1 0.045	4.1114	1.7584	2.338	0.01965	*
f2 4	-1.9226	1.3620	-1.412	0.15852	
f2 5	1.8604	1.3620	1.366	0.17240	
f3 0.4	-0.3416	1.1121	-0.307	0.75884	
f4 0	1.2721	1.1121	1.144	0.25306	
f1 0.045 :f2 4	5.4250	1.9262	2.816	0.00499	**
f1 0.045 :f2 5	6.0484	1.9262	3.140	0.00176	**
f1 0.045 :f3 0.4	0.1274	1.5727	0.081	0.93547	
f1 0.045 :f4 0	2.8621	1.5727	1.820	0.06921	.

TABLE 4.10: Coefficients modèle linéaire $c_2 \sim f_1 + f_2 + f_3 + f_4 + f_1 * f_2 + f_1 * f_3 + f_1 * f_4$

a. Signif. codes : 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

ne sont pas influencés de manière significative par le facteur f_4 .

Impact de la vitesse de propagation sur les autres impacts

Le tableau 4.10 fournit un exemple où l’on peut voir l’influence des valeurs des facteurs f_2 à f_4 sur la valeur du critère c_2 (perte de production élevage) en fonction de leurs interactions avec le facteur f_1 (taux de transmission de la maladie). On peut y voir notamment que les facteurs f_2 (période d’échantillonnage) et f_4 (connaissance du paramètre γ) ont un effet significativement plus fort sur le critère c_2 lorsque le facteur f_1 est à 0.045 (vitesse de propagation rapide). Ce résultat est encourageant car il montre que l’on peut traiter par l’expérience des questions de type 3 présentées dans la section 4.3 (qui questionnent l’influence de certaines caractéristiques de A sur les performances d’utilisation du modèle A_B^*).

4.6 Enseignements sur le système T_{uc}

4.6.1 Réponses aux questions sur le système T_{uc}

Les résultats du plan d’expérience illustrent bien les effets de la boucle de rétroaction entre le modèle de SPOC (ρ_o^*), le modèle de cadre expérimental utilisé pour le modèle SIS (A^{**} cadré et ρ_m^*) et le modèle de système épidémiologique (A_C^*). Ils nous permettent notamment de comparer deux méthodes de calibration du modèle SIS (en utilisant CE_{C1} ou CE_{C2}) non pas uniquement sur la base de la capacité prédictive du modèle SIS calibré, mais sur la base des trajectoires des sorties du système épidémiologique (trajectoires de prévalence, de ratio de contrôle etc.). On a pu voir par exemple que pour les modalités testées, selon les critères définis, il était toujours préférable de diminuer le pourcentage d’individus échantillonnés. Notons que ce résultat ne nécessite pas de chiffrage monétaire des coûts ou de manques à gagner.

Pour étendre les résultats d’expérience obtenus pour le modèle T_{uc}^* (présentés dans la section 4.5) au système T_{uc} , il faudrait réaliser une analyse de la sensibilité des critères que l’on observe aux paramètres utilisés pour la paramétrisation du modèle

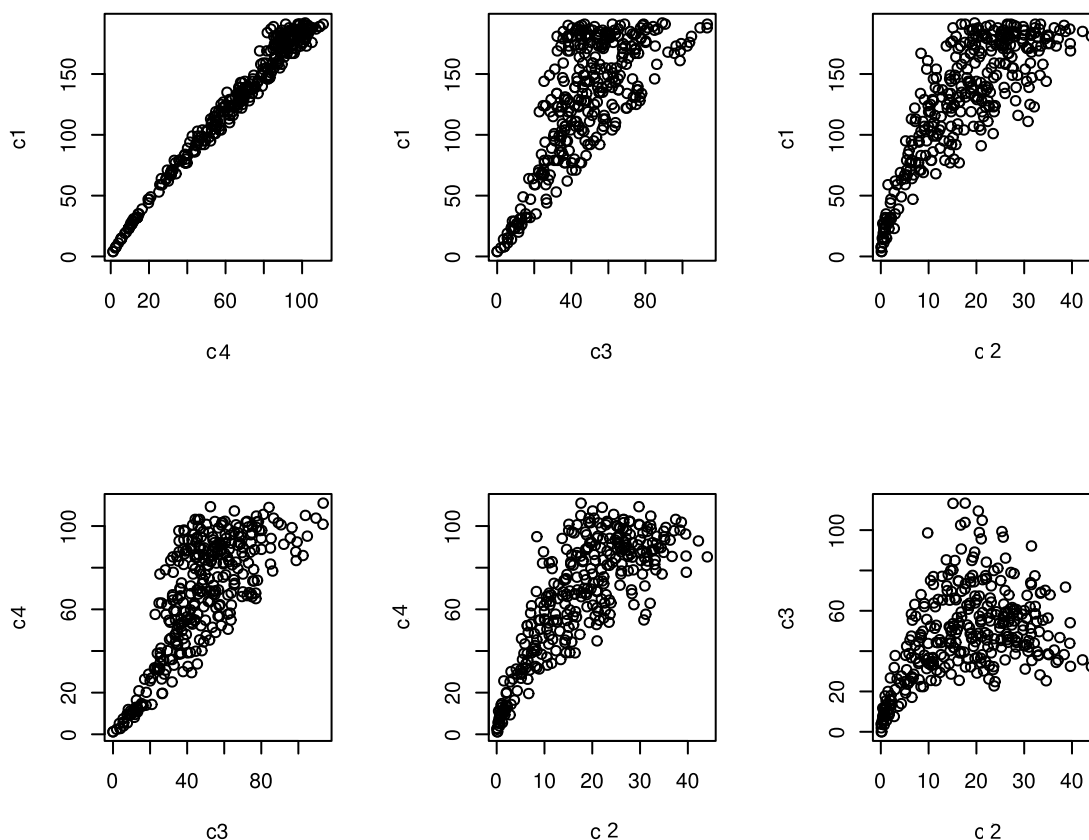


FIGURE 4.10: Relations entre les critères

T_{uc}^* (listés dans le tableau 4.4). Cette analyse n'a pas encore été réalisée mais une des suites de cette thèse sera d'élargir le plan d'expérience illustratif présenté section 4.5 (en ajoutant des modalités aux facteurs continus) et d'utiliser une des nombreuses méthodes d'analyse de sensibilité qui existent [STCR04] pour tester la dépendance des résultats obtenus à certaines valeurs des paramètres présentés dans le tableau 4.4.

4.6.2 Nouvelles questions sur T_{uc}

Critères d'évaluation de la Triade

Dans la section 4.5, nous avons considéré les différences entre les moyennes des critères calculés pour des expériences réalisées avec différentes modalités. La moyenne a été choisie pour les analyses statistiques parce qu'elle illustre suffisamment les réponses que l'on peut apporter aux questions posées dans la section 4.3 qui sont des questions générales (Est-ce que la surveillance a un impact ? Peut-on quantifier cet impact ? etc.). Dans une optique d'aide à la décision, il est possible de considérer d'autres indicateurs que les résultats moyens. On peut par exemple considérer des quantiles particuliers si l'on veut discuter de la robustesse de différentes

conceptions du SPOC. Le tableau 4.11 donne les quantiles à 95% du critère c_2 (perte de production élevage) pour des ensembles d'expériences réalisées avec différentes valeurs de modalités des facteurs f_1 et f_2 . Les valeurs apparaissant dans ce tableau traduisent la robustesse d'une configuration du SPOC dans le sens où la première ligne du tableau, par exemple, traduit le fait que dans 95% des expériences réalisées avec une période d'échantillonnage courte (3 ut), le critère c_2 (perte de production élevage) a pu être maintenu au dessous de 32 %.ut. On peut voir que, pour les valeurs de modalité choisies, la modification du facteur f_2 joue sur la robustesse du SPOC alors que le facteur f_3 n'a pas d'influence.

Facteur	Modalité	Quantile à 95%
f_2	Courte	32 %.ut
	Moyenne	39 %.ut
	Longue	43 %.ut
f_3	Précise	39%.ut
	Peu précise	39%.ut

TABLE 4.11: Quantile à 95% du critère c_2 (perte de production élevage) en fonction des facteurs f_2 et f_3

Autres Questions génériques sur les épidémies contrôlées

Le modèle T_{uc}^* peut être utilisé pour répondre à d'autres questions sur le système T_{uc} . Une question de type 3 (qui questionnent l'influence de certaines caractéristiques de A sur les performances d'utilisation du modèle A_B^*), que l'on peut adresser avec le modèle T_{uc}^* porte sur le fait que le modèle SIS ne prend pas en compte la structure des interactions entre les unités épidémiologiques. En considérant le paramètre « graphe » du modèle de système épidémiologique comme un facteur, il est possible d'identifier des caractéristiques du réseau d'infection pour lesquelles le modèle SIS permet au SPOC de contrôler plus ou moins bien la maladie.

4.7 Conclusion

La triade T_{uc}' nous a montré qu'il était possible de répondre à des questions sur une triade de Minsky (T_{uc}) en construisant un modèle T_{uc}^* et en expérimentant ce modèle.

Chapitre 5

Discussion

RÉSUMÉ

Trois outils génériques ont été développés au cours de cette thèse. Le cadre conceptuel de triade de triade T', la spécification formelle du modèle T* et une implémentation informatique du modèle T* dans la plate-forme Virtual Laboratory Environment (VLE). Ces outils sont des prototypes nécessaires à l'argumentation de notre thèse. Ils ont été construits dans un objectif de généralité mais ils restent orientés par notre cas d'application. Certaines limites et perspectives sont donc déjà identifiables pour poursuivre le développement de ces trois outils. L'objectif de cet ensemble d'outils est d'amener *in fine* de nouvelles perspectives pour la modélisation et la simulation des systèmes dynamiques. Suivant les types de systèmes considérés, ces outils pourront être plus ou moins bien adaptés. Certains résultats sont déjà clairement identifiables en ce qui concerne la surveillance épidémiologique, nous pouvons les décliner pour le cas des systèmes socio-écologiques, et des systèmes dynamiques en général.

5.1 Introduction : une contribution en trois outils

L'objectif de cette thèse, annoncé dans l'introduction, était d'apporter des éléments réponses à la problématique suivante : « Quelle validité des modèles en l'absence d'expérience reproductible ? ». Pour remplir cet objectif, nous avons du développer des méthodes et des outils qui devraient pouvoir être ré-utilisés par la communauté scientifique mais gagneraient à être améliorés. Ce chapitre de discussion est dédié à la présentation des limites et des perspectives que nous avons identifiées pour un ensemble de trois outils développés au cours de cette thèse.

Le premier est un outil conceptuel. Il s'agit du cadre conceptuel de triade de triade que nous avons noté T'. Il nous a servi à expliciter notre problématique dans le chapitre 1. Il est illustré par la figure 1.7, page 21. Il nous donne une base conceptuelle pour formuler les questions réflexives sur l'activité de modélisation et de simulation.

Le deuxième est un outil formel, il s'agit du modèle générique de triade que nous avons noté T*. Le chapitre 3 est dédié à sa présentation. La figure 3.3 page 56 illustre la structure générale de ce modèle générique.

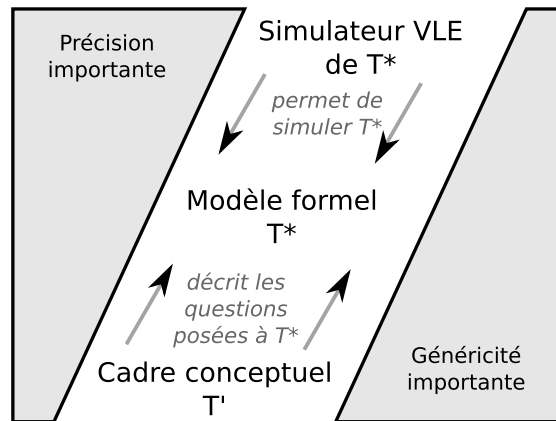


FIGURE 5.1: Une contribution à trois étage. Pour apporter des éléments réponses à la problématique : « Quelle validité des modèles en l'absence d'expérience reproductible ? », trois outils sont proposés : Le cadre conceptuel T' , le modèle formel T^* , et une implémentation d'un simulateur de T^* dans la plate-forme VLE. Le cadre conceptuel nomme les entités et relations que l'on considère. Il doit rester générique. Le modèle formel spécifie plus précisément les comportements dynamiques de ces relations et de ces entités. Certaines parties peuvent être génériques tandis que d'autres vont dépendre de la triade T modélisée. Enfin le simulateur simule le comportement spécifié par un modèle T^* donné. On attend qu'il respecte le comportement spécifié avec une très grande précision.

Le troisième est un outil informatique, il s'agit de l'implémentation (sous la plate-forme Virtual Laboratory Environment (VLE)) d'une instance de modèle de triade T^* modulaire dont certains composants sont génériques. Les détails de cette implémentation ont été peu développés dans le corps du manuscrit mais il nous semble important de le citer.

Les réponses que ces trois outils permettent d'apporter à notre problématique de recherche sont présentées dans notre chapitre de conclusion. Le cadre conceptuel T' , très abstrait, permet d'exprimer toute sorte de question portant sur l'usage d'une représentation quelconque, d'un modèle. À ce stade on ne parle pas encore de système dynamique mais uniquement des relations entre réalité, observateur et modèle. Le modèle formel T^* permet de spécifier l'entité T^* pour un cas particulier de triade de triade, à savoir lorsque l'entité triade T est identifiable à un réseau de systèmes dynamiques. L'utilisation du modèle formel T^* réduit ainsi le champs d'application du cadre conceptuel T' . En contrepartie, il permet de spécifier intégralement le modèle des interactions entre les différents composants de la triade T et permet donc de préciser les questions que l'on se pose sur T . Enfin, la construction d'un simulateur de T^* justifie la définition de questions sous la forme d'un cadre expérimental dans lequel va être simulé le modèle T^* . La figure 5.1 résume l'articulation entre les trois outils.

5.2 Des outils à enrichir

5.2.1 Le cadre conceptuel T'

Des modèles pour comprendre ou pour décider

Nous avons distingué trois types d'utilisation de modèles dans l'exposé de notre problématique : la synthèse, la prédiction ou l'explication. À ces types d'utilisations correspondent des relations différentes entre l'observateur et la réalité (ρ_o) et entre l'observateur et le modèle (ρ_m). Par exemple, dans le cas d'une utilisation d'explication, la relation ρ_o va consister à observer la réalité et la relation ρ_m va consister à induire le modèle. Dans le cas d'une utilisation de prédiction au contraire, la relation ρ_m va consister à déduire du modèle les scénarios probables et la relation ρ_o va consister à s'adapter à la réalité ou éventuellement à la modifier pour assurer le futur préféré. Tout au long du manuscrit nous nous sommes focalisés sur les cas où le modèle A_B^* de la triade T étudiée était utilisé dans un but de synthèse ou de prédiction. Nous pensons en effet que la construction du modèle T* et de son simulateur est plus aisée pour ces types de triades. Le chapitre d'état de l'art (chapitre 2) décrit les interactions entre les entités de la triade T pour ces cas de figures.

Dans l'exemple d'illustration donné chapitre 4, l'utilisation de modèle que l'on décrit vise davantage à synthétiser la dynamique du système épidémiologique qu'à l'expliquer (la dynamique de l'épidémie est synthétisée par l'indicateur R_0). Dans ce cas d'application, nous avons pu modéliser le changement d'attitude de l'observateur de Minsky entraîné par l'utilisation du modèle SIS. Nous l'avons modélisé par un modèle de processus de décision basé sur la valeur de l'indicateur R_0 . D'une manière générale, une utilisation de modèle à but synthétique est réalisée dans un objectif de prise de décision. Mieux encore, les synthèses fournies par ces modèles sont souvent des indicateurs choisis pour leur capacité à aider à la décision. Dans le cas du modèle SIS par exemple, l'indicateur R_0 est représentatif de la proportion d'individus infectés à l'équilibre. Si $R_0 \leq 1$, la proportion d'individus infectés à l'équilibre est nulle (cf. chapitre d'état de l'art section 2.5.1). Nous pensons que ce cas d'application en épidémiologie illustre bien le fait que ce type d'utilisation des modèles (la synthèse), se prête bien à l'utilisation de l'ensemble des trois outils : cadre T', modèle T* et simulateur VLE. Pour ce cas précis, le rôle de l'observateur B se modélise facilement sous la forme d'un processus de décision, la relation ρ_m sous la forme d'un processus d'expérimentation de modèle et la relation ρ_o sous la forme d'un SPOC. De plus, le processus de décision, qui caractérise l'influence du modèle (et du processus expérimental) sur le SPOC, découle très clairement du processus expérimental. Le modèle T* peut ainsi être construit.

On a vu dans le chapitre 4 que le modèle DS-DEVS T*, spécifié de manière de manière incomplète, permet de construire aisément le simulateur d'un tel modèle.

On a également vu dans ce même chapitre comment le cadre conceptuel T' se prête à l'expression des questions que l'on a sur T et des réponses que l'on attend de T*.

Dans le cas d'une utilisation de modèle à but prédictif, les résultats obtenus par le processus expérimental sont des prédictions, données pour un certain nombre de scénarios. Nous pensons que pour ce cas d'utilisation, la théorie de la décision dans l'incertain (cf. section 2.6.1 page 39) nous permet de modéliser le lien entre

les prédictions obtenues et les décisions prises sur le SPOC. Les prédictions doivent pour cela être associées à une mesure de leur probabilité ou possibilité d'occurrence. Cela est souvent prévu lors de l'utilisation de modèles dans un but prédictif. Le modèle formel T^* et le simulateur VLE que nous proposons peut donc également être utilisé pour les triades comportant des modèles à buts prédictifs, à condition de mobiliser la théorie de la décision dans l'incertain. Les modèles décisionnels fournis par cette théorie devront pour cela être implémentés dans le modèle de processus de décision apparaissant dans le modèle T^* .

Dans le cas d'une utilisation de modèle à but explicatif, un travail important reste à réaliser. Ce travail doit commencer au niveau du cadre conceptuel T' . Il serait nécessaire de préciser la nature des entités A , B et A^* et des relations ρ_m et ρ_o pour ces types d'utilisation. Pour l'instant, ces entités et relations sont définies de manière très abstraites. Nous avons essayé de les expliciter dans le chapitre 2 mais un travail plus approfondi serait nécessaire si l'on veut pouvoir déduire le modèle formel T^* du cadre conceptuel T' dans le cas d'une triade T qui décrirait une utilisation de modèle à but explicatif.

Nous avons vu dans le chapitre 2 que dans le contexte thématique qui nous intéresse (les socio-éco-systèmes), les méthodes de construction de modèles « pour comprendre » étaient encore en cours de formalisation. Rappelons le protocole ODD qui vise à fournir un standard pour la spécification de modèles écologiques destinés à la recherche scientifique [GBD⁺10]. Nous avons entrepris des travaux visant à réunir la grille de lecture d'utilisation des modèles fournie par la TMS (qui mobilise les concepts de morphismes, validation, cadre expérimental, etc.) avec celle fournie par le protocole ODD (qui mobilise les concepts d'émergence, d'adaptation, etc.) [DBQsp]. Ces travaux donnent des pistes pour la définition d'une méthodologie canonique de l'utilisation de modèle pour la compréhension des systèmes complexes.

Une fois le cadre conceptuel T' enrichi, il sera éventuellement nécessaire d'adapter le modèle formel T^* . On a pu voir dans les paragraphes qui précèdent que les utilisations de modèles pour la synthèse ou la prédiction pouvaient être reliées de manière rationnelle à la prise de décision. C'est ce qui nous permet de modéliser la relation entre le processus expérimental et le SPOC. Dans le cas d'une modélisation pour « comprendre », la rétroaction du processus expérimental sur le système cible est plus difficile à définir.

5.2.2 Le modèle formel T^*

Généricité du Modèle de Processus Expérimental

Considérer les modèles A_B^* comme des MSD nous a amené à formaliser un Modèle de Processus Expérimental (MPE) générique. Nous pensons qu'il s'agit d'une contribution à part entière indépendamment du reste du modèle T^* . Le MPE a d'ailleurs fait l'objet d'un article co-écrit dans le cadre cette thèse mais appliqué à un autre cadre d'utilisation que celui de la modélisation de la triade de Minsky. L'exemple d'utilisation qui avait été choisi portait sur la modélisation multi-échelles. Cet exemple n'est pas nécessaire dans l'argumentation de cette thèse mais il illustre bien l'utilisation du MPE et nous l'avons donc mis en annexe (cf. annexe F).

Notre objectif en construisant le MPE était de développer une méthode qui permette de concevoir et d'implémenter la méthode de simulation encapsulée, com-

prise comme une expérience virtuelle réalisée à l'intérieur d'une expérience virtuelle. Nous voulons proposer un cadre pour spécifier comment les expériences virtuelles sont choisies et exécutées de manière itérative et dynamique. Notre motivation est de fournir une méthodologie générique pour spécifier et donc implémenter le couplage entre un tel processus d'expérimentation de modèle et n'importe quel MSD. Nous proposons ainsi un cadre générique pour inclure ce processus d'expérimentation de modèles dans une simulation. Pour être opérationnel, ce cadre générique est conçu sous la forme d'un modèle de système dynamique qu'est le MPE.

Nous pensons qu'à terme, le MPE doit permettre de spécifier des modèles composés de systèmes dynamiques qui n'évoluent pas dans le même monde en spécifiant intégralement des couplages asynchrones. L'intérêt est évident quand il s'agit d'étudier l'utilisation de modèles de systèmes dynamiques comme nous l'avons fait dans le cadre de la modélisation de la triade de Minsky. Nous pensons que d'autres cas d'application existent, tel celui de la modélisation multi-échelles qui nécessite de faire interagir des systèmes définis à des échelles spatio-temporelles différentes (cf. annexe F).

Notons qu'un Modèle de Processus Expérimental ainsi construit peut être utilisé seul pour implémenter un processus expérimental, tel qu'un algorithme d'optimisation par simulation, des analyses de sensibilité, des études de viabilités de systèmes etc. Cependant l'intérêt du MPE n'est pas là. La plupart des environnements de modélisation et de simulation proposent déjà des fonctionnalités pour que le moteur de simulation récupère les valeurs des paramètres et des états initiaux depuis un espace de travail global, que les simulations puissent être lancées depuis cet espace de travail, et que les résultats de simulation puissent être rapatriés dans l'espace de travail. Donnons comme exemple la plateforme PowerDEVS qui est interfacée avec Scilab [BK11] ou la plateforme VLE (celle que l'on utilise) qui est interfacée avec le logiciel de traitement statistique « R » [QDR09]. Ces fonctionnalités sont très utiles et très utilisées mais elles n'offrent pas de formalisme pour coupler l'exécution d'une simulation avec un autre modèle de système dynamique. C'est la raison d'être du MPE. L'intérêt du MPE réside dans sa relation avec le modèle dans lequel il est lui-même inclus. Le modèle couplé résultant du couplage entre le MPE et un autre modèle de système dynamique spécifié dans un formalisme conforme à DEVS, peut lui-même être spécifié en DEVS.

Différents types de modèles A_B^*

Le modèle T^* tel que nous l'avons spécifié dans le chapitre 3 se concentre sur les cas où le modèle A_B^* est un modèle de simulation. Pour ces cas précis, la TMS fournit un cadre d'utilisation des modèles qui passe par la formulation de cadre expérimentaux et que nous avons formalisé sous la forme du modèle de processus expérimental. Notons toutefois que l'on peut s'intéresser à des triades dans lesquelles le modèle A_B^* n'est pas un modèle de système dynamique. Une formalisation de l'utilisation de modèles statistiques serait notamment intéressante.

Remettons nous dans le contexte de la surveillance épidémiologique pour appuyer ces propos. Dans notre exemple d'application, des modèles statistiques sont mobilisés dans la relation ρ_o de l'observateur à l'objet. La surveillance épidémiologique est en effet basée sur un échantillonnage qui nécessite un minimum d'analyses statistiques. Ces analyses statistiques sont elles-mêmes basées sur des modèles, tels

que des distributions connues par exemple. Dans l'objectif d'étudier l'utilisation de ces modèles statistiques, ils peuvent être considérés comme des entités A_B^* . Cela nécessiterait une formalisation différente du modèle T^* qui pourrait intégrer des classes de modèles statistiques usuelles en tant qu'entité A_B^* ainsi que des méthodes d'utilisation telles que le choix de modèle par des critères de performance type AIC comme relation ρ_m^* . Un tel développement gagnerait à être réalisé en parallèle à un enrichissement du cadre conceptuel T .

5.2.3 L'implémentation modulaire de T^* dans VLE

Un des avantages du formalisme DS-DEVS est qu'il permet une implémentation modulaire. Ainsi, une fois que la dynamique d'un simulateur atomique est implémentée, il est possible de la réutiliser à volonté dans différents modèles composés. De plus, de part la propriété de fermeture sous couplage de DEVS, un modèle composé peut lui-même être intégré dans un autre modèle composé.

Nous avons développé tous nos modèles dans l'environnement de développement VLE qui offre de nombreuses fonctionnalités pour le développement de modèles dans le formalisme DS-DEVS. L'environnement VLE propose ainsi un développement sous forme de paquet et la possibilité d'inclure un modèle composé, préalablement défini, à l'intérieur d'un autre modèle. L'ensemble des simulateurs qui ont été codés pour la construction du modèle T^* sont donc réutilisables indépendamment et un paquet indépendant a été développé pour le modèle de processus expérimental.

5.3 Des outils à utiliser

5.3.1 Application à la surveillance épidémiologique en santé animale

Vers un système de surveillance basé sur la simulation

Le système T_{uc} tel que nous l'avons décrit dans le chapitre 4 n'existe pas aujourd'hui de manière aussi formalisée. Il existe bien des interactions entre les équipes de modélisation et les décideurs responsables de la surveillance et du contrôle des maladies infectieuses mais ces interactions ne sont pas encore formalisées au point où un protocole d'échange d'information et des règles de décisions soient définis de manière aussi précise que nous les avons décrits dans le système T_{uc} .

Cependant, le fait que l'on puisse expérimenter ce type de système en utilisant des modèles de triade du type de T_{uc}^* devrait permettre de mettre en place de nouveaux types de réseaux de surveillance basés sur la simulation. Il faudrait pour cela utiliser un modèle A^{**} spatialisé. Rappelons que nous calibrons l'état et les paramètres du modèle A^{**} en fonction des cas détectés par la surveillance. Le modèle de la triade ainsi constituée nous permettrait d'évaluer la capacité d'un modèle A^{**} donné et d'une méthode de calibration ρ_m^* donnée à identifier les états possibles d'un modèle de système épidémiologique A_C^* . Grâce à cela, il serait possible d'évaluer théoriquement un système de surveillance qui utiliserait un modèle et une méthode de calibration pour reconstituer l'état du système épidémiologique surveillé et qui

utiliserait les résultats de simulations obtenus pour répartir spatialement les efforts de surveillance sur le système épidémiologique.

Notons que l'utilisation des modèles en épidémiologie est étudiée (cf. section 1.3.2) et des études récentes maintiennent la communauté scientifique au courant de la manière dont les modèles sont utilisés pour réguler le système de contrôle [SPH11], le système de surveillance [WPL11], ou dans la gestion générale d'une crise épidémique ou d'une maladie particulière [MDTH11, SBdJ11]. Ces études peuvent nous fournir une base de données pour valider un modèle de système triade du type du système T_{uc} .

5.3.2 Application aux systèmes sociologiques et écologiques

Une application particulière du cadre de Marvin Minsky

Marvin Minsky évoquait déjà dans son article de 1968, un observateur M , qui disposerait d'un modèle W^* du monde W , et qui voudrait utiliser un modèle W^{**} qui le représenterait lui même (l'observateur) et qui contiendrait donc une entité M^* . Minsky expliquait alors que la relation entre le modèle (W^*) et (W^{**}) pouvait être difficile à cerner. Pour reprendre ses termes : « *the reader may be anxious, at this point, for more details about the relation between W^* and W^{**}* ». Si l'observateur s'inclut dans le modèle W^{**} , il doit nécessairement y inclure le modèle qu'il utilise (W^*). Or, si l'observateur est conscient de ce niveau de réflexivité, le modèle W^* qu'il utilise contient déjà le modèle W^{**} . Minsky en déduit tout d'abord que le terme « contenir » n'est pas suffisamment sophistiqué pour décrire la relation entre W^* et W^{**} . Mais il en déduit également que le terme de « modèle » n'est pas suffisant pour étudier techniquement les relations entre les modèles. Il explique qu'une analyse adéquate aura besoin d'idées beaucoup plus avancées sur la représentation symbolique des structures de traitement d'informations. Nous sommes d'accord avec ce point en ce qui concerne l'utilisation que Minsky fait des modèles dans son article. Dans le cas où le modèle sert à expliciter la manière de penser d'une « créature », nous manquons encore aujourd'hui de connaissance sur la manière dont ces modèles de penser se créent et sur les briques qui les composent.

Toutefois, dans le contexte thématique des SES, les modèles, la manière et les objectifs pour lesquels ils sont construits, ainsi que l'utilisation qu'on en fait, sont connus. Les briques qui manquent à la psychologie pour décrire les processus cognitifs sont disponibles en écologie pour décrire des démarches de modélisation ou des modes de gestion des écosystèmes, elles font l'objet du chapitre « d'État de l'art » de la thèse (cf. chapitre 2). Dans ce cas précis, nous connaissons les mécanismes ayant menés au modèle A_B^* et nous pouvons expliciter la rétroaction que l'utilisation de A_B^* va avoir sur A .

L'exemple que nous avons donné en épidémiologie est certes schématique mais nous pensons qu'il illustre tout de même bien ces propos. Le modèle A_B^* de T_{uc} doit être choisi parmi l'ensemble des modèles épidémiologiques existant, qui sont formalisés, caractérisés (modèles spatialisés ou non, stochastiques ou non, etc.), puis calibré sur la base d'observations que l'on peut décrire (un système de surveillance). De même, les mesures prises sur la base des informations tirées du modèle A_B^* peuvent être explicitées. C'est le fait que l'on puisse expliciter ces relations qui nous permet de construire un modèle T_{uc}^* suffisamment précis pour que l'on puisse

simuler son comportement. On peut ainsi dans notre cas expérimenter l'objet T_{uc}^* en tant que modèle de système dynamique.

Modélisation et prise de décision

Dans sa présentation à la conférence ESSA, Alan Kirman (GREQAM, Université Paul Cézanne, EHESS) soulignait les problèmes liés à l'utilisation de modèles pour les sciences économiques et leurs implication dans la récente crise financière [Kir11]. Il donne plusieurs citations de responsables « économiques » que nous reprenons ici.

When the crisis came, the serious limitations of existing economic and financial models immediately became apparent. Arbitrage broke down in many market segments, as markets froze and market participants were gripped by panic. Macro models failed to predict the crisis and seemed incapable of explaining what was happening to the economy in a convincing manner. As a policy-maker during the crisis, I found the available models of limited help. In fact, I would go further : in the face of the crisis, we felt abandoned by conventional tools. In the absence of clear guidance from existing analytical frameworks, policy-makers had to place particular reliance on our experience. Judgement and experience inevitably played a key role.

Trichet (2010) Gouverneur de la Banque Centrale Européenne.

et

But there is also a strong belief, which I share, that bad or rather oversimplistic and overconfident economics helped create the crisis. There was a dominant conventional wisdom that markets were always rational and self-equilibrating, that market completion by itself could ensure economic efficiency and stability, and that financial innovation and increased trading activity were therefore axiomatically beneficial.

Adair Turner, Head of the U.K. Financial Services Authority

La question se pose donc clairement aujourd'hui de l'impact des activités scientifiques de modélisation et de simulation sur la société civile. Que les responsables que nous avons cité aient raison ou tort, il est nécessaire d'explicitier les relations entre les modèles et les systèmes que ces modèles permettent de gérer. Notre travail devrait aider à se rendre compte de ces impacts en étudiant les triades de Minsky de façon rationnelle et empirique.

Le cadre de modélisation de triades de Minsky que nous proposons devrait pouvoir être utilisées dans un but pédagogique de manière à donner une certaine forme d'expérience à des décideurs utilisateurs de modèles. Il devrait permettre d'identifier les situations dans lesquelles les modèles disponibles ne permettent plus de prendre de décision sur le système cible. Il devrait également montrer comment ce sont parfois les modèles utilisés pour prendre des décisions qui ont menés à ces situations car ils ne permettaient pas de les imaginer.

5.3.3 Application à l'étude des systèmes dynamiques en général

Niveaux d'abstraction des triades de Minsky

La question initiale qui nous a menée à construire ces outils portait sur la validation des modèles. Les outils que nous avons développés doivent nous permettre de formuler plus précisément des questions liées à la validité des modèles mais ils n'excluent pas le problème de la validation des modèles. En effet, la question se pose de la validation du modèle T^* . Cette question n'est pas traitée dans cette thèse. Nous considérons simplement que le modèle T^* est suffisamment pertinent pour traiter les questions que l'on se pose sur l'objet T . C'est à dire que les simplifications qui sont faites entre le modèle A^{**} et le modèle A_C^* sont bien représentatives des simplifications qui nous intéressent entre le modèle A_B^* et l'objet A , que les relations ρ_m et ρ_o sont correctement représentées par les modèles ρ_m^* et ρ_o^* , etc.

Les apports des outils développés ne portent en fait pas sur la validation des modèles mais plutôt sur leur évaluation dans un contexte d'utilisation donné. Nous avons dit que la relation entre le modèle A^* et le modèle A^{**} avait été identifiée par M. Minsky comme une question sensible et importante. Dans notre cas, le passage de l'entité A_B^* à l'entité A^{**} fait partie d'une montée en abstraction qui porte sur l'ensemble de la triade de Minsky. L'entité A^{**} est un modèle de l'entité A_B^* qui est elle même un modèle du système cible réel A . La relation entre A_B^* et A^{**} est donc claire, aucun des deux n'est contenu dans l'autre, A_B^* est contenu dans T et A^{**} est contenu dans T^* qui est un modèle de T . Ce que nous étudions avec le modèle T^* est l'interrelation entre le modèle, l'observateur et l'objet modélisé que l'on observe dans T dans le cadre d'un certain nombre de questions que l'on se pose sur le système T et non sur le système A . Se poser la question de la pertinence de cette démarche requiert en fait de monter en abstraction.

On pourrait construire désormais un modèle T'^* de la triade T' et on formerait ainsi une nouvelle triade T'' qui nous permettrait de nous poser des questions sur notre utilisation de T^* . Nous pouvons continuer ainsi indéfiniment. La figure 5.2 illustre cette montée en abstraction qui ne consiste plus à construire des modèles de modèles mais des triades de triades. Nous appellerons par la suite le niveau d'abstraction 0 une triade faisant intervenir un objet de la réalité, le niveau d'abstraction 1 le niveau de la triade T' , le niveau 2 est le niveau de la triade T'' qui fait intervenir T' et T'^* , etc. Nous discutons l'intérêt d'une telle montée en abstraction dans la section suivante.

Montée dans les niveaux d'abstraction

Comme nous l'avons expliqué dans la section 5.2.1, certains développements sont nécessaires pour étendre notre cadre de modélisation et de simulation de la triade de Minsky aux triades faisant intervenir des modèles explicatifs. Nous supposons dans ce paragraphe que ces développements sont faits. La question se pose alors de savoir ce que la modélisation et l'expérimentation de la triade de Minsky nous permet d'apprendre au final sur le système cible A . Cette question est légèrement décalée puisque le modèle de la triade nous permet de nous poser des questions sur le système triade et non sur un système cible A qui en serait indépendant. Cependant,

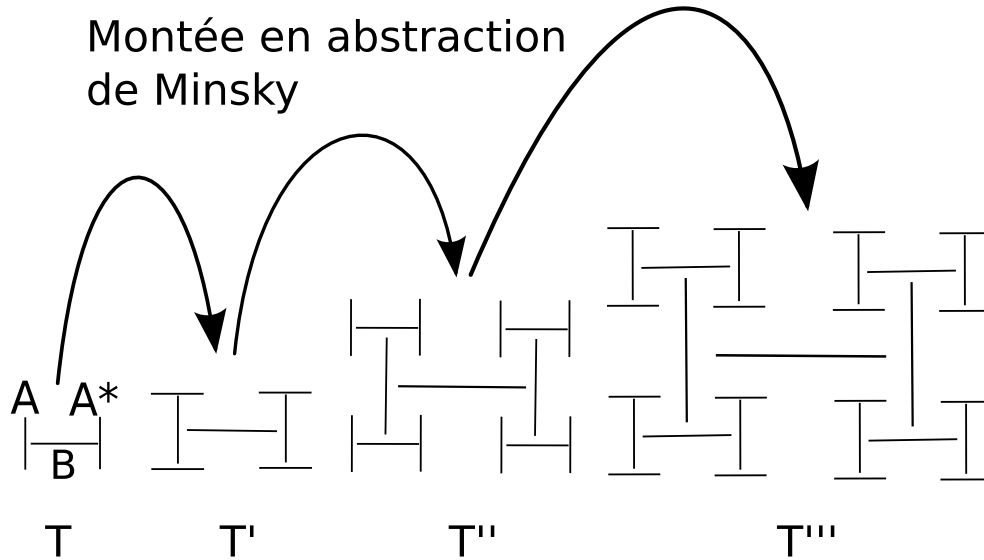


FIGURE 5.2: Montée en abstraction de Minsky. Comme le montre la triade T' , la modélisation d'une triade de Minsky produit une nouvelle triade de Minsky. Il est possible de construire ainsi une infinité de « triade de triades » à des niveaux de plus en plus abstraits. Dans cette figure, une triade de Minsky est représentée sous la forme suivante : les représentations de l'objet et du modèle sont orientées selon le même axe (horizontal ou vertical) et reliés par la représentation de l'observateur qui est un segment de droite perpendiculaire aux représentations de l'objet et du modèle. Les représentations de l'objet et du modèle sont définies récursivement comme étant soit un segment de droite dans le cas où l'objet est un objet du monde réel, soit, dans le cas contraire, une représentation de triade de Minsky telle qu'elle est décrite précédemment. De gauche à droite, la première triade est une triade dont l'objet fait partie du monde réel, la seconde triade est une triade T' .

le système cible étant modélisé dans le modèle de la triade (composant A_C^* de T^*), le modèle de la triade nous permet également de nous poser des questions sur A . En fait, le fait de nous interroger sur ce que le modèle de la triade de Minsky nous permet d'apprendre sur A , revient à nous interroger sur ce que le modèle de la triade de Minsky nous apprend sur le système triade qu'il modélise. En restant dans la démarche constructiviste que nous avons décrite dans la section 5.3.3, nous serions tentés de monter d'un niveau d'abstraction (cf. section précédente) en construisant un modèle T'^* .

Notre chapitre d'état de l'art a pour objectif de montrer que la modélisation de la triade de Minsky de niveau d'abstraction 0 est possible dans le cas de la modélisation des systèmes épidémiologiques, sociologiques ou écologiques. Nous pensons malheureusement que l'état des connaissances ne permet pas aujourd'hui de modéliser une triade de niveau d'abstraction 1. Cela reviendrait à modéliser et simuler l'activité de réfléchir sur les modèles que nous utilisons. En attendant que l'utilisation du cadre que nous proposons se généralise pour le niveau 1 et puisse ainsi être analysée, modélisée et simulée, les réflexions ayant lieu au niveau d'abstraction 1 devons nous contenter d'analyses philosophiques plus classiques mobilisant éventuel-

lement des modèles conceptuels ou des expériences de pensées mais qui sortent tout de même du cadre d'application de la théorie de la modélisation et de la simulation.

Cette démarche n'a théoriquement pas de fin. La figure 5.3 illustre l'amorce d'un cadre de réflexion infini qui engloberait tous les niveaux d'abstractions.

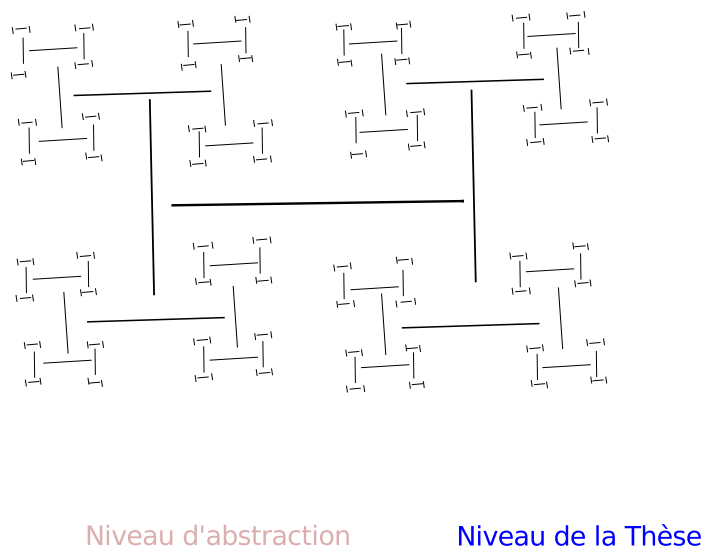


FIGURE 5.3: Cadre infini de la connaissance de Minsky. Cette figure illustre l'amorce d'un cadre de réflexion infini qui engloberait tous les niveaux d'abstractions de Minsky (cf. figure 5.2).

Constructivisme et épistémologie expérimentale

De manière plus théorique, la combinaison des trois outils : le cadre conceptuel T' , le modèle formel T^* et le simulateur VLE, nous permet d'étudier la relation entre une réalité et l'utilisation d'un modèle au sein de cette réalité. Ces outils nous permettent plus précisément de mener cette étude de manière empirique, en réalisant des simulations. Dans de nombreux cas, les « erreurs » de prédiction observées sont induites par des simplifications introduites consciemment dans le modèle par le modélisateur. Cependant, une chose est de savoir qu'on introduit une erreur lorsque l'on ne considère pas les évolutions des seconds moments d'une distribution (cf. annexe E section 1), une autre est de comparer visuellement les trajectoires prédites et les trajectoires réelles (cf. annexe E figure 4), et encore une autre de pouvoir comparer les conséquences d'utiliser un modèle plutôt qu'un autre sur la trajectoire du système modélisé (cf. section 4.5). Nous pensons que ces trois points de vues sont importants et complémentaires pour aller jusqu'au bout de notre activité de modélisation, notamment pour faire prendre conscience aux décideurs éventuels des biais induits par la modélisation.

Dans l'exposé de notre problématique, nous avons justifié une motivation pragmatique de construire un cadre de modélisation et de simulation de la triade de Minsky. Ce cadre nous semble en effet nécessaire pour étudier et contrôler de manière rationnelle des systèmes sur lesquels on ne peut réaliser d'expériences. Cependant,

nous avons également une motivation paradigmatique qui consistait à simplement offrir la possibilité de construire et d'expérimenter un système triade artificiel. En utilisant T^* , il est possible d'avoir recours à tous les outils et concepts des sciences expérimentales. En construisant T^* on apprend sur la relation complexe entre le modèle et le système modélisé dans T .

Un meilleur modèle pour de meilleures décisions ?

En modélisant et en simulant la triade de Minsky, on peut évaluer la pertinence de l'utilisation du modèle au regard de la trajectoire qu'il nous permet d'imposer au système cible. Nous avons pu le voir dans notre chapitre d'application sur le modèle T_{uc}^* . Cependant, le système cible que nous asservissons dans le modèle T^* reste un modèle. Il s'agit du modèle A_C^* . On ne peut pas savoir avec certitude si un modèle A^{**} qui nous permet d'asservir correctement le système A_C^* nous permettra d'asservir l'objet A avec le même succès. En revanche, on peut distinguer des types de systèmes A_C^* et des types de modèles A^{**} et étudier si certains types de modèles A^{**} sont plus ou moins bien adaptés à la gestion de certains types de systèmes A_C^* . Dans ce cas, on pourra en déduire quels types de modèles A_B^* sont les mieux adaptés à la gestion d'un certain type de systèmes A . Comme exemple de types de modèles nous pouvons penser à des modèles spatialisés versus non spatialisés pour des dynamiques écologiques ou à des modèles individus-centrés versus des modèles agrégés pour des dynamiques quelconques.

Optimisation de méthodes de calibration de modèles avec observation invasive et politiques d'observations dynamique

Nous caractérisons « d'invasive » une observation qui entraîne nécessairement une modification du système observé et de « dynamique » une politique d'observation qui peut évoluer au cours du temps. Nous pensons que la construction d'un simulateur de T^* peut permettre d'optimiser des méthodes de calibration de modèles basées sur des politiques d'observations qui sont dynamiques et dont les observations sont invasives.

Nous nous intéressons particulièrement aux méthodes de calibration basées sur des méta-heuristiques d'optimisation par simulation. Optimiser une telle heuristique revient à optimiser une méthode d'optimisation et on parle parfois de méta-optimisation. Dans notre exemple nous avons utilisé la méta-heuristique d'optimisation par essaim de particules. Nous avons paramétré cette méta-heuristique en utilisant des valeurs générales trouvées dans la littérature (cf. section 4.4.3). Grâce au simulateur du modèle T^* , les valeurs de paramètres de l'algorithme OEP (paramètres ω , ϕ_g , ϕ_p et N du tableau D.1) peuvent être calibrés par simulation en cherchant par exemple les valeurs de paramètres qui permettent un nombre de simulations minimal, tout en convergeant vers les bonnes solutions (c'est à dire les solutions qui permettent de *fitter* au mieux les données d'observation).

Ensuite, dans le même processus d'optimisation, il est possible de faire varier les paramètres de conception du système de surveillance. Il est ainsi possible d'optimiser également le système d'observation, en parallèle avec l'algorithme de calibration de modèle.

Nous n'avons pas choisi notre exemple sur le cas d'une maladie à contrôle obligatoire. L'exemple sans contrôle obligatoire nous semblait suffisant pour une illustration d'utilisation du cadre. Cependant, comme nous le disions dans l'exposé de notre problématique, c'est un cas particulièrement intéressant à étudier avec un modèle de triade de Minsky. Dans ce cas, les observations réalisées sur le système cible peuvent le modifier directement en entraînant une rétroaction directe du SPOC, nous parlons dans ce cas d'observation « invasive » puisque l'observation a un impact sur l'état du système observé. En représentant les effets de ces rétroactions dans le cadre expérimental de calibration (CE_C présenté dans l'annexe D) il est possible de les prendre en compte dans la calibration du modèle. Le simulateur de T^* permet donc d'implémenter des méthodes d'optimisation de méthodes de calibration de modèles qui comprennent une politique d'observation dynamique avec des observations invasives.

Notons que l'on peut trouver dans la littérature récente des travaux visant à optimiser les « modèles » utilisés dans ce type de circonstances [YM10]. Il serait intéressant de faire un pont entre ces méthodes et les outils développés dans le cadre de la modélisation et de la simulation de la triade de Minsky.

Chapitre 6

Conclusion

RÉSUMÉ

Nous nous sommes intéressés dans cette thèse à la modélisation des Socio-Éco-Système (SES) inexpérimentables. Pour ce type de systèmes, il est impossible de valider un modèle de manière classique, c'est à dire en comparant les sorties du modèle avec les sorties du système pour un cadre expérimental donné. La thèse a consisté à proposer un cadre méthodologique pour évaluer la pertinence de l'utilisation de modèles pour la gestion de tels systèmes. Ce cadre est composé de trois outils basés sur la notion de triade de Minsky (noté T) qui considère explicitement l'objet modélisé A, l'observateur B et le modèle de l'objet A*. Le premier outil est le cadre conceptuel T' de triade de triade. Le second est une spécification formelle générique du modèle de triade T*. Le dernier est une implémentation informatique d'un simulateur du modèle T*. Ces trois outils, s'ils sont utilisés avec précaution, pourront apporter des réponses dans les domaines de la surveillance épidémiologique en santé animale, de la modélisation des SES et de la M&S des systèmes dynamiques en général.

6.1 Un cadre méthodologique pour l'évaluation des modèles

Nous nous sommes intéressés dans cette thèse à la modélisation des Socio-Éco-Système (SES) sur lesquels on ne peut pas reproduire d'expérience. Pour ce type de systèmes, il est impossible de valider un modèle de manière classique, c'est à dire en comparant statistiquement les sorties du modèle avec les sorties du système pour un cadre expérimental donné. Notre thèse a consisté à proposer et défendre un cadre méthodologique pour évaluer la pertinence de l'utilisation de modèles dans ces cas où ils ne peuvent pas être validés de manière classique. Nous n'avons pas défini de barème spécifique pour quantifier la pertinence d'un modèle par rapport à un autre mais nous proposons un cadre conceptuel pour définir une grille d'évaluation qui se focalise sur le contexte d'utilisation des modèles. Nous avons noté T' ce cadre

conceptuel. Ce cadre conceptuel nous a mené à la construction d'un modèle formel T^* qui constitue le cœur de notre proposition. Le modèle formel T^* nous permet de spécifier formellement un modèle et son contexte d'utilisation. Cela nous permet de construire un simulateur de T^* et ainsi d'expérimenter T^* .

Nous avons appliqué le cadre méthodologique proposé à un exemple en surveillance épidémiologique en santé animale (noté T_{uc}) qui montre que les trois outils, cadre conceptuel T' , modèle formel T^* et simulateur de T^* permettent la mise au point d'une grille d'évaluation quantitative des modèles de systèmes sur lesquels on ne peut pas reproduire d'expérience.

6.1.1 Le cadre conceptuel T'

Nous sommes parti de la définition de modèle que donne Marvin Minsky : Pour un observateur B, un objet A^* est un modèle d'un objet A, dans la mesure où B peut utiliser A^* pour répondre à une question qui l'intéresse sur A. Les objets, A, B et A^* constituent la triade de Minsky T. Notre proposition a été de considérer cette triade comme un système dynamique et de construire un modèle de ce système que l'on note T^* . Nous avons pour cela explicité les relations que l'observateur B entretient avec l'objet A (notée ρ_o) et avec le modèle A^* (notée ρ_m) dans le cas de la modélisation et de la gestion des SES. Le modèle générique T^* que nous proposons permet de modéliser les entités A, B et A^* , et les relations ρ_o et ρ_m . L'objectif de la thèse était de montrer qu'il était possible et intéressant de construire et de simuler le modèle T^* pour le cas où l'objet A était un SES sur lequel on ne peut pas reproduire d'expérience. Le chapitre 2 a expliqué pourquoi il était possible de construire le modèle T^* dans le cas des SES, le chapitre 3 a montré comment construire ce modèle et le chapitre 4 a montré sur un exemple comment utiliser le modèle T^* .

Nous voulons souligner ici que la démarche de construction et d'utilisation du modèle T^* a été intégrée dans un cadre conceptuel que nous avons commencé à ébaucher dans cette thèse et que nous avons noté T' (la présentation de ce cadre fait l'objet du chapitre 1).

En construisant le modèle T^* nous nous positionnons en tant qu'observateur qui se pose des questions sur T et construit le modèle T^* pour répondre à ces questions. Nous avons noté C cet observateur. le cadre conceptuel T' est la nouvelle triade T-C- T^* où l'objet est la triade T, le modèle est T^* et l'observateur est C. Comme nous l'avons développé dans notre chapitre de discussion (section 5.2.1), le cadre conceptuel T' gagnerait à être enrichi et développé. La relation entre C et T, définit les questions que nous pouvons nous poser sur une triade T. Nous avons noté ρ_{o_c} cette relation mais nous ne l'avons pas détaillée de manière générale. De la même manière, nous avons noté ρ_{m_c} la relation entre C et T^* . Cette dernière relation correspond à l'utilisation d'un système dynamique et nous pouvons utiliser la notion de cadre expérimental développé par la TMS puisque le modèle T^* est un Modèle de Système Dynamique (MSD). Notre chapitre d'application (le chapitre 4) donne un exemple de triade T' et donne ainsi des exemples de relations ρ_{o_c} et ρ_{m_c} .

6.1.2 La spécification formelle du modèle T^*

Le cœur de cette thèse est la proposition d'un modèle générique de triade de Minsky. Nous l'avons noté T^* . Le chapitre 3 est dédié à une spécification formelle du modèle T^* qui utilise le formalisme DS-DEVS. Le modèle formel obtenu situe chaque modèle d'entité et de relation de la triade T^* dans un modèle DS-DEVS. Dans ce modèle, chacune de ces entités ou relation est représentée par un modèle de système dynamique ou un ensemble de MSD couplés. Le formalisme DS-DEVS permet une décomposition hiérarchique de telle manière que certaines entités ou relations peuvent être représentées comme des systèmes composés de sous-systèmes composants. Parmi les modèles composants, certains sont génériques et ont été intégralement spécifiés au plus haut niveau de spécification (l'ensemble de leur dynamique est défini), d'autres dépendent naturellement de la triade modélisée et ont été spécifiés au niveau de spécification le plus bas (seuls les ports d'entrées/sorties sont définis).

Le composant correspondant au modèle ρ_m^* de la relation de l'observateur au modèle est remarquable. Ce composant permet de faire la liaison entre un modèle (le modèle A_C^*) et un modèle de modèle (le modèle A^{**}). L'objectif d'une spécification formelle du modèle T^* nous a amené à spécifier formellement ce lien. Nous nous sommes basé pour cela sur la notion de cadre expérimental développée dans la TMS qui vise à spécifier formellement l'utilisation qui est faite d'un modèle. À partir de cette notion, nous avons développé un Modèle de Processus Expérimental générique qui permet de spécifier le caractère dynamique de l'usage d'un modèle et donc de coupler un MSD au processus d'utilisation d'un autre MSD. Comme nous l'avons développé dans notre chapitre de discussion (section 5.2.2), nous pensons que la spécification que nous avons proposé pour le modèle de processus expérimental est une bonne piste pour la spécification de couplage asynchrone en DEVS.

6.2 Limites et perspectives d'utilisation

Les voies qui nous semblent avoir été ouvertes dans le domaine de l'utilisation de la modélisation en épidémiologie, pour la gestion et la compréhension des SES en général, et plus théoriquement pour la M&S des systèmes dynamiques sont présentées dans la section 5.3 de notre discussion. Nous résumons ici ces limites et perspectives pour conclure ce manuscrit.

6.2.1 Un cadre d'évaluation réflexif et relatif

Le premier point à rappeler est que l'ensemble des outils que nous proposons ne peut assurer la validation d'un modèle dans l'absolu. En effet, notre cadre méthodologique utilise un modèle T^* dont la validité est laissée à l'appréciation du modélisateur. Or les résultats obtenus sur T^* ne sont transposables à T uniquement dans la mesure où T^* est un modèle valide de T . La validation des modèles ρ_o^* , ρ_m^* , B^* et A^{**} peut éventuellement être établie de manière classique. La validation du modèle A_C^* ne peut en revanche pas être obtenue pour le cas qui nous intéresse puisqu'il est impossible de reproduire d'expérience sur l'objet A . Le fait que l'utilisation de A^{**} donne de bons résultats au sein de la triade T^* , n'assure pas que le

modèle A_B^* donnera de bons résultats au sein de la triade T. Il est donc nécessaire de délimiter avec précaution les questions auxquelles on peut répondre sur la triade T en utilisant un modèle T^* . C'est seulement au regard de ces questions que l'on peut ou non considérer le modèle T^* comme « valide ». De manière générale, nous pensons que cette méthodologie et les outils associés devrait permettre d'identifier des typologies de modèles plus ou moins bien adaptés à des typologies de SES pour certains contexte d'utilisation (cf. section 5.3.3 page 114). C'est à dire que les questions auxquelles on pourra répondre se formuleront de la manière suivante : « Si le système A respecte telle hypothèse, il est plus judicieux d'utiliser ce type de modèles (spatialisé, non spatialisé, ..) pour obtenir ce type de résultat (contrôle, compréhension, prédiction ...) ». L'avantage d'utiliser un modèle de triade est que cela permet d'exprimer explicitement ce type de questions réflexives sur notre activité de M&S et d'y répondre de manière empirique et quantitative.

6.2.2 Des pistes concrètes pour la surveillance épidémiologique et la gestion des SES

L'utilisation de l'ensemble des outils proposés a été illustrée par un exemple d'application sur la surveillance épidémiologique en santé animale. Le chapitre 4 décrit cette application. Il montre comment la simulation d'un modèle T^* permet d'évaluer quantitativement une situation dans laquelle un système de surveillance épidémiologique et un MSD sont utilisés pour la gestion d'une épidémie. Nous expliquons dans notre chapitre de discussion pourquoi la possibilité d'évaluer théoriquement ce type de situation ouvre la porte à la conception de systèmes de surveillance basés sur la simulation (cf. section 5.3.1).

Les systèmes épidémiologiques font partie des SES au sujet desquels de nombreux modèles ont été développés. Bien que les liens entre l'épidémiologie théorique et la prise de décision en santé publique ne soit pas formalisée, une certaine culture de modélisation accompagne les décisions prises quant-à la gestion des épidémies. Pour cette raison, notre application nous semble bien adaptée pour illustrer l'utilisation du cadre conceptuel T' et du modèle T^* . Toutefois, nous tenons à faire remarquer que d'autres types de SES devraient se prêter tout aussi bien à ce type d'étude. Comme liste non-exhaustive on pourrait citer la gestion des pêcheries, la gestion des dynamiques forestières ou le domaine de la finance.

Un dernier point plus général reste à ajouter. Nous avons déjà expliqué comment la simulation du modèle T^* nous permet d'évaluer le modèle d'un système dynamique sur lequel on ne peut pas reproduire d'expérience. Il est en fait possible d'aller plus loin en disant que cela nous permet également d'évaluer le modèle d'un système dynamique que l'on ne pourrait observer sans le perturber. De plus, étant donné que l'ensemble de la triade est simulé, la politique d'observation est évaluée en même temps que le modèle qu'elle permet de construire. Le fait de pouvoir évaluer la combinaison entre la politique d'observation et le modèle devrait nous permettre d'optimiser des démarches de modélisation (ou de calibration de modèles) de systèmes dynamiques que l'on ne peut observer sans perturber. Cette piste est développée dans la section 5.3.3 (page 114) et explicitée sur le cas d'un exemple en épidémiologie.

6.2.3 Un point de vue constructiviste

Pour conclure, j'aimerais revenir sur l'épigraphe choisi pour ce manuscrit. Il s'agit d'une citation de Paul Valéry parlant de Léonard de Vinci¹.

Il ne concevait pas de savoir véritable auquel ne correspondit pas quelque pouvoir d'action. Créer, construire étaient pour lui inséparable de connaître et comprendre.

P. Valéry, Préface aux Cahiers de Léonard de Vinci, 1942

À l'instar des constructivistes, il me semble que l'on ne peut apprendre ni comprendre sans construire. Cette thèse est soutenue dans la discipline informatique et mobilise des concepts très abstraits tels que les modèles de modèles. Toutefois, la majeure partie du temps passé sur ce travail de thèse est davantage assimilable à des activités d'ingénierie et de conception qu'à la lecture de documents théoriques ou à la réflexion méditative. Cela peut difficilement transparaître dans un manuscrit de thèse et c'est pourquoi je tiens à le souligner ici. La construction de simulateurs DEVS passe tout d'abord par des étapes de conception des sous-modèles souvent en collaboration avec des experts des processus modélisés. Ensuite, de nombreuses techniques et outils sont mobilisés pour programmer et utiliser les simulateurs. Citons dans notre cas l'environnement VLE comme Interface de Programmation d'Application DEVS, le langage de programmation C++ comme langage de programmation informatique, le compilateur, le logiciel R et les outils statistiques associés pour créer les interfaces de visualisation ou de traitement statistiques des résultats de simulations, le logiciel de gestionnaire de version Git pour assurer un suivi du code informatique, etc... La mise au point du modèle T_{uc}^* et de son simulateur fut ainsi pour moi, de fait, indissociable d'une démarche de conception et de construction à tâtons d'un programme opérationnel et intéressant.

La démonstration qu'il est possible aujourd'hui de construire des simulateurs de modèles T^* me semble ainsi être le résultat le plus intéressant de cette thèse. J'espère que, de la construction de nombreux simulateurs de modèles T^* , émergera une meilleure compréhension de la triade T en général.

1. Épigraphe du chapitre intitulé Sur les fondements épistémologiques de la science de la cognition : contribution de la systémique aux constructivismes par J.L. Le Moigne dans [AMT⁺91]

Bibliographie

- [ADD08] Edouard AMOUROUX, Stéphanie DESVAUX et Alexis DROGOUL : Towards virtual epidemiology : An agent-based approach to the modeling of h5n1 propagation and persistence in north-vietnam. *In* Springer Berlin / HEIDELBERG, éditeur : *Intelligent Agents and Multi-Agent Systems*, volume Volume 5357/2008, pages 26–33, 2008.
- [AM79] Roy M. ANDERSON et Robert M. MAY : Population biology of infectious diseases : Part i and ii. *Nature*, 280:361–367, 1979.
- [AMT⁺91] E. ANDREEWSKY, J.-L. Le MOIGNE, J.-C. TABARY, J.-L. VULLIERME, R. VALLÉ, B. BOUCHON-MEUNIER et J.B. GRIZE : *Systémique et Cognition*. Afcet Systèmes. Dunod, 1991.
- [AR86] K. J. ARROW et H. RAYNAUD : *Social choice and Multicriteria decision making*. MIT Press, Cambridge, 1986.
- [Aum07] Graig A. AUMANN : A methodology for developing simulation models of complex systems. *Ecological Modelling*, 202:385–396, 2007.
- [BAD⁺99] P. BONJOUR, J.L. ANGOT, Barbara DUFOUR, M. SANAA et B. TOMA : une méthode quantitative d'évaluation des services vétérinaires dans le cadre de l'analyse de risque à l'importation d'animaux vivants. *Épidémiologie et santé animale*, 36:39–63, 1999.
- [Bar96] Fernando J. BARROS : Dynamic structure discret event system specification : Formalism, abstract simulators and applications. *Winter Simulation*, 13(1):35–46, 1996.
- [Bar97] Fernando J. BARROS : Modeling formalisms for dynamic structure systems. *ACM Trans. Model. Comput. Simul.*, 7:501–515, October 1997.
- [Bar03] Fernando J. BARROS : Dynamic structure multiparadigm modeling and simulation. *ACM Trans. Model. Comput. Simul.*, 13:259–275, July 2003.
- [BBPSV05] Marc BARTHÉLEMY, Alain BARRAT, Romualdo PASTOR-SATORRAS et Alessandro VESPIGNANI : Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Journal of Theoretical Biology*, 235(2):275 – 288, 2005.
- [BD11] B. BONTÉ et Raphaël DUBOZ : A modular simulation tool to help designing epidemics surveillance : Work in progress. Poster présenté à la conférence ICAHS, 17-20 Mai, Lyon, France, 2011.
- [BDQM09] Bruno BONTÉ, Raphaël DUBOZ, Gauthier QUESNEL et Jean-Pierre MÜLLER : Recursive simulation and experimental frame for multiscale

- simulation. In *SCSC'09 : Summer Computer Simulation Conference*, July 13th-16th 2009.
- [BK11] Federico BERGERO et Ernesto KOFMAN : Powerdevs : a tool for hybrid system modeling and real-time simulation. *SIMULATION*, 87(1-2): 113–132, 2011.
- [Bol99] Benjamin M. BOLKER : Analytic models for the patchy spread of plant disease. *Bulletin of Mathematical Biology*, 61(5):849 – 874, 1999.
- [Bom09] Pierre BOMMEL : *Définition d'un cadre méthodologique pour la conception de modèles multi-agents adaptée à la gestion des ressources renouvelables*. Thèse de doctorat, Université Montpellier II, mai 2009.
- [Bou94] Francois BOUSQUET : *Des Milieux, des Poissons, des Hommes : Etude par Simulations Multi-Agents. Le cas de la Pêche dans le Delta Central du Niger*. Thèse de doctorat, Université Claude Bernard, Lyon 1, mars 1994.
- [BP04] F BOUSQUET et C Le PAGE : Multi-agent simulations and ecosystem management : a review. *Ecological Modelling*, 176(3-4):313–332, 2004.
- [BPPTss] Bruno BONTÉ, Éric PENOT, Christophe. Le PAGE et Jean-François TOURRAND : *modélisation des exploitations agricoles avec Olympe (Seconde édition)*, chapitre Liaison d'un outil de modélisation d'exploitations agricoles (Olympe) avec une plateforme de modélisation multi-agents (Cormas). Quae, 2011 (under press).
- [BPT05] Bruno BONTÉ, Éric PENOT et Jean-François TOURRAND : Coupling the farming system modelling tool 'olympé' with the multi-agent-system software system 'cormas' to understand the use of resources in complex agricultural systems. In *ECMS 2005 : 19th European Conference on Modelling and Simulation*, June 1st-4th 2005.
- [BQMDew] Bruno BONTÉ, Gauthier QUESNEL, Jean-Pierre MÜLLER et Raphaël DUBOZ : Modelling of the experimentation process : Application to scale-transfer. *Simulation : Transactions of the Society for Modeling and Simulation International*, 21(2):187–202, Under Review.
- [Can09] R.M. CANNON : Estimating prevalence from targeted surveillance. *Preventive Veterinary Medicine*, In Press, Corrected Proof:–, 2009.
- [Car08] O. CARTON : *Langages formels, calculabilité et complexité*. Vuibert, 2008.
- [CB05] I. CHADÈS et B. BOUTEILLER : Solving multiagents mdps : Forest management example. In *Proceedings of MODSIM-05*, 2005.
- [Cha82] A.F. CHALMERS : *Qu'est-ce que la science ? Popper, Khun, Lakatos, Feyerabend. 2de édition*. La Découverte(Ed)., 1982.
- [CW09] Rachid CHREYH et Gabriel WAINER : Cd++ repository : an internet based searchable database of devs models and their experimental frames. In *Proceedings of the 2009 Spring Simulation Multiconference, SpringSim '09*, pages 159 :1–159 :8, San Diego, CA, USA, 2009. Society for Computer Simulation International.

- [DBDG09] Raphael DUBOZ, Bruno BONTE, Stephanie DESVAUX et Vladimir GROSBOIS : Simulation of avian influenza propagation in the poultry production network of the north vietnam. Poster presentation at the EPIDEMICS conference, Athens, Greece, 2-4 December, 2009.
- [DBQsp] Raphaël DUBOZ, Bruno BONTÉ et Gauthier QUESNEL : Vers une spécification des modèles de simulation de systèmes complexes. *CFP : Modélisation informatique et mathématique des systèmes complexes : avancées méthodologiques. Numéro Spécial de la revue Studia Informatica Universalis*, 2011sp.
- [DDS⁺04] B. DURAND, M. A. DUBOIS, P. SABATIER, D. CALAVAS, C. DUCROT et A. Van DE WIELLE : Multiscale modelling of scrapie epidemiology : Ii. geographical level : hierarchical transfer of the herd model to the regional disease spread. *Ecological Modelling*, 179(4):515–531, 2004.
- [DF11] Stéphanie DESVAUX et Muriel FIGUIÉ : Formal and informal surveillance systems : how to build bridges? *In Proceedings of the first ICAHS Conference*, pages 1942–1948, 2011.
- [DH07] Barbara DUFOUR et Pascal HENDRIKX : *Surveillance épidémiologique en santé nimale - 2e édition*. Quae, 2007.
- [DL00] U. DIECKMANN et R. LAW : *The geometry of ecological interactions : simplifying spatial complexity*. mbridge University Press, Cambridge, UK, 2000.
- [DMS10] Ian DOHOO, Wayne MARTIN et Henrik STRYHN : *Veterinary Epidemiologic Research, 2nd Edition*. VER Inc., 2010.
- [DP01] Didier DUBOIS et Henri PRADE : Possibility theory, probability theory and multiple-valued logics : A clarification. *Annals of Mathematics and Artificial Intelligence*, 32:35–66, August 2001.
- [Dro93] Alexis DROGOUL : *De la simulation multi-agent à la résolution collective de problèmes. Une étude de l'émergence de structures d'organisation dans les systèmes multi-agents*. Thèse de doctorat, Université Paris VI, 1993.
- [Dub04] Raphaël DUBOZ : *Intégration de modèles hétérogènes pour la modélisation et la simulation de systèmes complexes : application au transfert d'échelles en écologie marine*. Thèse de doctorat, Université du Littoral Côte d'Opale, mars 2004.
- [Dur94] R. DURRETT : The importance of being discrete (and spatial). *Theoretical Population Biology*, 46(3):363–394, December 1994.
- [ERE09] K.T.D. EAMES, J.M. READ et W.J. EDMUNDS : Epidemic prediction and control in weighted networks. *Epidemics*, 1(1):70–76, 2009.
- [EU09] Robert E. et ULANOWICZ : The dual nature of ecosystem dynamics. *Ecological Modelling*, 220(16):1886 – 1892, 2009.
- [FAESL09] Jenny FRÖSSLING, Estelle Carina Constance AGREN, Lena ELIASSON-SELLING et Susanna Sternberg LEWERIN : Probability of freedom from disease after the first detection and eradication of prrs in sweden : scenario-tree modelling of the surveillance system. *Prev Vet Med*, 91(2-4):137–45, 2009.

-
- [Fer95] Jacques FERBER : *Les systèmes multi-agents. Vers une intelligence collective*. InterEditions, Paris, 1995.
 - [FG01] J. FILIPE et G. GIBSON : Comparing approximations to spatio-temporal models for epidemics with local spread. *Bulletin of Mathematical Biology*, 63:603–624, 2001. 10.1006/bulm.2001.0234.
 - [FK99] M. A. FUENTES et M. N. KUPERMAN : Cellular automata and epidemiological models with spatial dependence. *Physica A : Statistical and Theoretical Physics*, 267(3-4):471 – 486, 1999.
 - [FMB10] Stefano FAROLFI, Jean-Pierre MÜLLER et Bruno BONTÉ : An iterative construction of multi-agent models to represent water supply and demand dynamics at the catchment level. *Environmental Modelling & Software*, 25(10):1130 – 1148, 2010.
 - [Fol06] C. FOLKE : Resilience : The emergence of a perspective for social ecological systems analyses. *Global Environmental Change*, 16:253–267, 2006.
 - [Fou98] Emmanuel FOUQUET : *Dictionnaire Hachette encyclopédique*. Hachette, 1998.
 - [Fra04] Alain FRANC : Metapopulation dynamics as a contact process on a graph. *Ecological Complexity*, 1(1):49–63, 2004.
 - [Gal91] G.C. GALLOPIN : Human dimension of global change : linking the local and the global processes. *International Social Science Journal*, 130:707–718, 1991.
 - [GBA⁺10] Carole A. GOBLE, Jiten BHAGAT, Sergejs ALEKSEJEVS, Don CRUICKSHANK, Danius MICHAELIDES, David NEWMAN, Mark BORKUM, Sean BECHHOFFER, Marco ROOS, Peter LI et David DE ROURE : myExperiment : a repository and social network for the sharing of bioinformatics workflows. *Nucleic Acids Research*, 38(suppl 2):W677–W682, 2010.
 - [GBB⁺06] Volker GRIMM, Uta BERGER, Finn BASTIANSEN, Sigrunn ELIASSEN, Vincent GINOT, Jarl GISKE, John GOSS-CUSTARD, Tamara GRAND, Simone K. HEINZ, Geir HUSE, Andreas HUTH, Jane U. JEPSEN, Christian JORGENSEN, Wolf M. MOOIJ, Birgit MÜLLER, Guy PE’ER, Cyril PIOUS, Steven F. RAILSBACK, Andrew M. ROBBINS, Martha M. ROBBINS, Eva ROSSMANITH, Nadja RÜGER, Espen STRAND, Sami SOUSSI, Richard A. STILLMAN, Rune VABO, Ute VISSER et Donald L. DEANGELIS : A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1-2):115 – 126, 2006.
 - [GBD⁺10] V. GRIMM, U BERGER, DL DEANGELIS, G POLHILL, J GISKE et SF RAILSBACK : The odd protocol : a review and first update. *Ecological Modelling*, 221:2760–2768, 2010.
 - [GP94] J.B. GREENBERG et T.K. PARK : Political ecology. *Journal of Political Ecology*, 1(1), 1994.
 - [GS05] John B. GILMER et Frederick J. SULLIVAN : Issues in event analysis for recursive simulation. In M. E. KUHL, N. M. STEIGER, F. B.

- ARMSTRONG et J. A. Joines and EDS., éditeurs : *Proceedings of the 2005 Winter Simulation Conference*, pages 1234–1241. The Society for Computer Simulation, july 2005.
- [HK10] Thomas HOUSE et Matt J. KEELING : Insights from unifying modern approximations to infections on networks. *Journal of The Royal Society Interface*, 2010.
- [Höh07] Michael HÖHLE : An r package for the monitoring of infectious diseases. *Computational Statistics*, volume 22(4):571–582, 2007.
- [HPH09] Michael HÖHLE, Michaela PAUL et Leonhard HELD : Statistical approaches to the monitoring and surveillance of infectious diseases for veterinary public health. *Preventive Veterinary Medicine*, In Press, Corrected Proof:–, 2009.
- [HXP⁺09] Thomas D. HARWOOD, Xiangming XU, Marco PAUTASSO, Mike J. JEGGER et Michael W. SHAW : Epidemiological risk assessment using linked network and grid based modelling : *Phytophthora ramorum* and *phytophthora kernoviae* in the uk. *Ecological Modelling*, In Press, Corrected Proof:–, 2009.
- [KE95] J. KENNEDY et R. EBERHART : Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks*, pages 1942–1948. IEEE Press, 1995.
- [KE05] Matt J KEELING et Ken T.D EAMES : Networks and epidemic models. *Journal of The Royal Society Interface*, 2(4):295–307, 2005.
- [Kee05] Matt KEELING : The implications of network structure for epidemic dynamics. *Theoretical Population Biology*, 67(1):1 – 8, 2005.
- [Kir11] Alan KIRMAN : Agent based models and the crisis. In *Proceedings of ESSA Conference, Montpellier*, 2011.
- [KJ01] Ernesto KOFMAN et Sergio JUNCO : Quantized state systems. a DEVS approach for continuous systems simulation. 18:123–132, 2001.
- [Kli85] G.J. KLIR : *Architecture of Systems Complexity*. Saunders, New York, 1985.
- [KR07] M.J. KEELING et P. ROHANI : *Modelling Infectious Diseases In Humans And Animals*. Princeton University Press, 2007.
- [KW99] M.N KUPERMAN et H.S WIO : Front propagation in epidemiological models with spatial dependence. *Physica A : Statistical Mechanics and its Applications*, 272(1-2):206 – 222, 1999.
- [KWS⁺01] Matt J. KEELING, Mark E. J. WOOLHOUSE, Darren J. SHAW, Louise MATTHEWS, Margo CHASE-TOPPING, Dan T. HAYDON, Stephen J. CORNELL, Jens KAPPEY, John WILESMITH et Bryan T. GRENFELL : Dynamics of the 2001 UK Foot and Mouth Epidemic : Stochastic Dispersal in a Heterogeneous Landscape. *Science*, 294(5543):813–817, 2001.
- [LD96] S. A. LEVIN et R. DURRETT : From individuals to epidemics. *Philosophical Transactions of the Royal Society of London. Series B : Biological Sciences*, 351(1347):1615–1621, 1996.

- [LMD03] R. LAW, D. MURRELL et U. DIECKMANN : Population growth in space and time : spatial logistic equations. *Ecology*, 84(1):252 – 262, 2003.
- [Lor63] Edward N. LORENZ : Deterministic non-periodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.
- [LPFL09] Catherine LINARD, Nicolas PONÇON, Didier FONTENILLE et Eric LAMBIN : Risk of malaria reemergence in southern france : Testing scenarios with a multiagent simulation model. *Ecohealth*, May 2009.
- [MDTH11] L.M. MANSLEY, A.I. DONALDSON, M.V. THRUSFIELD et N. HONHOLD : Destructive tension : mathematics versus experience, the progress and control of the 2001 foot and mouth disease epidemic in great britain. *Rev. sci. tech. Off. int. Epiz.*, 30(2):483–498, 2011.
- [Min65] Marvin MINSKY : Matter, mind and models. *In Proceedings of IFIP Congress*, pages 45–49, 1965.
- [MLPSV09] B. MARTÍNEZ-LÓPEZ, A.M. PEREZ et J.M. SÁNCHEZ-VIZCAÍNO : Social network analysis. review of general concepts and use in preventive veterinary medicine. *Transboundary and Emerging Diseases*, 56:109 – 120, 2009.
- [MRF⁺10] Alessandro MACCAGNAN, Mauro RIVA, Erika FELTRIN, Barbara SIMIONATI, Tullio VARDANEGA, Giorgio VALLE et Nicola CANNATA : Combining ontologies and workflows to design formal protocols for biological laboratories. *Automated Experimentation*, 2(1):3, 2010.
- [Mül04] Jean-Pierre MÜLLER : The mimosa generic modelling and simulation platform : The case of multi-agent systems. *In SCS. Proceedings of the 5th Workshop on Agent-Based Simulation*, pages 77–86. SCS, 2004.
- [NNGG89] Ernest NAGEL, James R. NEWMAN, Kurt GÖDEL et Jean-Yves GIRARD : *Le théorème de Gödel*. Éditions du Seuil, 1989.
- [Ost09] Elinor OSTROM : A general framework for analyzing sustainability of social-ecological systems. *Science*, 325(5939):419–422, 2009.
- [PCS⁺09] Sarah E. PERKINS, Francesca CAGNACCI, Anna STRADIOTTO, Daniele ARNOLDI et Peter J. HUDSON : Comparison of social networks derived from ecological data : implication for inferring infectious disease dynamics. *Journal of Animal Ecology*, 78:1015–1022, 2009.
- [Ped10] M.E.H. PEDERSEN : Good parameters for particle swarm optimization. Technical Report HL1001 (Hvass Laboratories), 2010.
- [PHT08] M. PAUL, L. HELD et A.M. TOSCHKE : Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine*, 27:6250–6267, 2008.
- [PJ08] Marco PAUTASSO et Michael J. JEGER : Epidemic threshold and network structure : The interplay of probability of transmission and of persistence in small-size directed networks. *Ecological Complexity*, 5(1):1 – 8, 2008.
- [QDR09] Gauthier QUESNEL, Raphaël DUBOZ et Éric RAMAT : The Virtual Laboratory Environment – An operational framework for multi-modelling, simulation and analysis of complex dynamical systems. *Simulation Modelling Practice and Theory*, 17:641–653, April 2009.

- [R D06] R DEVELOPMENT CORE TEAM : *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.
- [Sav54] L. SAVAGE : *The Foundations of Statistics*. Dover, 1954.
- [SBdJ11] J.A. STEGEMAN, A. BOUMA et M.C.M. de JONG : Epidemiological models to assist the management of highly pathogenic avian influenza. *Rev. sci. tech. Off. int. Epiz.*, 30(2):571–579, 2011.
- [SDD⁺04] P. SABATIER, B. DURAND, M.A. DUBOIS, C. DUCROT, D. CALAVAS et A. Van DE WIELLE : Multiscale modelling of scrapie epidemiology : I. herd level : a discrete model of disease transmission in a sheep flock. *Ecological Modelling*, 180(2-3):233–252, 2004.
- [Sin10] Alexander SINGER : External report reviewing the previous opinions of the panel on animal health and welfare concerning the application of quantitative tools, in the sequence of the current self mandate on "good practice in conducting scientific assessments in animal health using modelling". SCIENTIFIC REPORT submitted to EFSA (Question No EFSA-Q-2009-408), 2010.
- [SM09] P.H.T. SCHIMIT et L.H.A. MONTEIRO : On the basic reproduction number and the topological properties of the contact network : An epidemiological study in mainly locally connected cellular automata. *Ecological Modelling*, 220(7):1034 – 1042, 2009.
- [SPH11] C. SAEGERMAN, S.R. PORTER et M.-F. HUMBLET : The use of modelling to evaluate and adapt strategies for animal disease control. *Rev. sci. tech. Off. int. Epiz.*, 30(2):555–569, 2011.
- [SR05] Mark D.F. SHIRLEY et Steve P. RUSHTON : The impacts of network topology on disease spread. *Ecological Complexity*, 2(3):287 – 299, 2005.
- [SST11] Alexander SINGER, Mo SALMAN et Hans-Hermann THULKE : Reviewing model application to support animal health decision making. *Preventive Veterinary Medicine*, In Press, Corrected Proof:–, 2011.
- [STCR04] Andrea SALTELLI, Stefano TARANTOLA, Francesca CAMPOLONGO et Marco RATTO : *Sensitivity Analysis in Practice : A Guide to Assessing Scientific Models*. New York : John Wiley & Sons, 2004.
- [TDS⁺96] B. TOMA, B. DUFOUR, M. SANAA, J.J. BÉNET, P. ELLIS, F. MOUTOU et A. LOUZA : *Épidémiologie appliquée à la lutte collective contre les maladies transmissibles majeures*. Association pour l'étude de l'épidémiologie des maladies animales, Maisons-Alfort, France, 1996.
- [TDS⁺08] B. TOMA, B. DUFOUR, M. SANAA, J.J. BÉNET, A. SHAW et F. Moutou et A. LOUZÀ : *Epidémiologie appliquée à la lutte collective contre les maladies animales transmissibles majeures (2e édition)*. Broché, 2008.
- [TDZ08] Jean-Pierre TREUIL, Alexis DROGOUL et Jean-Daniel ZUCKER : *Modélisation et simulation à base d'agents - Exemples commentés, outils informatiques et questions théoriques*. 2008.
- [TGM⁺99] H. H. THULKE, V. GRIMM, M. S. MÜLLER, C. STAUBACH, L. TISCHENDORF, C. WISSEL et F. JELTSCH : From pattern to practice : a

- scaling-down strategy for spatially explicit modelling illustrated by the spread and control of rabies. *Ecological Modelling*, 117(2-3):179 – 202, 1999.
- [TM01] William W. TURECHEK et Laurence V. MADDEN : Effect of scale on plant disease incidence and heterogeneity in a spatial hierarchy. *Ecological Modelling*, 144(1):77 – 95, 2001.
- [Tra06] Muzy Alexandre TRAORE K. MAMADOU : Capturing the dual relationship between simulation models and their context. *Simulation Modelling Practice and Theory*, 14:126–142, 2006.
- [Tro02] Carl TROPPER : Guest editor's introduction : Parallel discrete-event simulation applications. *J. Parallel Distrib. Comput.*, 62(3):327–335, 2002.
- [Val91] R. VALLÉE : *Systémique et Cognition*, chapitre La cognition et le paradoxe de l'"observation - action". Afcet Systèmes. Dunod, 1991.
- [Var88] Francisco J. VARELA : *Les sciences cognitives*. Connaître. Seuil, 1988.
- [VDB⁺10] T. VERGNE, B. DUFOUR, B. BONTÉ, S. SAN, T. SOTHYRA, A. BOUCHOT, F. ROGER et V. GROBOIS : Simulation of cr methods as a new tool to assess animal disease surveillance : application to find outbreak reports in cambodia. Poster presentation SVPEM conference, Nantes, France, March 24-26, 2010.
- [VGB11] T. VERGNE, F. Goutard and D. Holland C. Bellet and F. Roger and V. GROBOIS et B. DUFOUR : Capture-recapture as a tool to assess animal disease surveillance : the example of foot-and-mouth disease in cambodia. *In Proceedings of the first ICAHS Conference*, pages 1942–1948, 2011.
- [VGS⁺10] T. VERGNE, V. GROBOIS, S. SAN, T. SOTHYRA, F. GOUTARD, B. BONTE, A. BOUCHOT, F. ROGER et B. DUFOUR : How to design a second source for an effective capture-recapture analysis? the example of foot-and-mouth disease in cambodia. *In AITVM conference (oral com.)*, August 13th-16th 2010.
- [VMU74] F.G. VARELA, H.R. MATURANA et R. URIBEA : Autopoiesis : The organization of living systems, its characterization and a model. *Bio-systems*, 5(4):187–196, 1974.
- [vNM47] J. von NEUMANN et O. MORGENSTERN : *Theory of Games and Economic Behavior*. Princeton University Press, 1947.
- [VTR93] Francisco J. VARELA, E. THOMPSON et E. ROSCH : *L'inscription corporelle de l'esprit*. Seuil, 1993.
- [Wal77] Bernard WALLISER : *Systèmes et modèles Introduction critique à l'analyse de systèmes*. Seuil, 1977.
- [Wei48] Norbert WEINER : *Cybernetics*. Paris, Hermann, 1948.
- [WEW09] Michael S. WILLIAMS, Eric D. EBEL et Scott J. WELLS : Population inferences from targeted sampling with uncertain epidemiologic information. *Preventive Veterinary Medicine*, 89(1-2):25 – 33, 2009.

- [Wik11] WIKIPÉDIA : Informatique — wikipédia, l'encyclopédie libre, 2011. [En ligne ; Page disponible le 16-octobre-2011].
- [WPL11] P. WILLEBERG, L.G. PAISLEY et P. LIND : Epidemiological models to support animal disease surveillance activities. *Rev. sci. tech. Off. int. Epiz.*, 30(2):603–614, 2011.
- [YM10] Levent YILMAZ et Bradley MITCHELL : *Activity-Based Modeling and Simulation*, chapitre Generative Multisimulation : Decision-Support under Uncertainty using Evolutionary Multimodels. Presses Universitaires Blaise-Pascal, 2010.
- [YTNK08] Takehisa YAMAMOTO, Toshiyuki TSUTSUI, Akiko NISHIGUCHI et Sota KOBAYASHI : Evaluation of surveillance strategies for bovine brucellosis in japan using a simulation model. *Preventive Veterinary Medicine*, 86(1-2):57–74, 2008.
- [Zad99] L. A. ZADEH : Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100(Supplement 1):9 – 34, 1999.
- [ZKP00] Bernard P. ZEIGLER, Tag Gon KIM et Herbert PRAEHOFFER : *Theory of modeling and simulation : Integrating Discrete Event and Continuous Complex Dynamic Systems*. Academic Press, 2000.

Annexes

Annexe A

The Virtual Laboratory Environment

A.1 Classe de modèle

Les classes de modèles de VLE se présentent comme des modèles composés. Une classe spécifie la structure du modèle composé (connexions entre ports), ainsi que les dynamiques et les conditions initiales de chaque modèle atomique composant. Les modèles de type "exécutif" ont accès aux classes de modèles définies dans un projet VLE et peuvent les utiliser pour créer des modèles composant qui aient la même spécification qu'une classe donnée. Naturellement, un identifiant différent doit être attribué pour chaque nouveau modèle créé à un niveau hiérarchique donné.

A.2 Graph translator

Une classe "GraphTranslator" est mise à disposition par l'API VLE. Elle permet de construire des structures de modèles couplés à partir d'une matrice d'adjacence et de la liste des classes des modèles à coupler. Un identifiant différent est attribué à chaque modèle composant (de manière manuelle par la spécification de la liste des identifiants ou de manière automatique). Il est également possible de spécifier différentes politiques de nommages des ports de connexions entre modèles. La spécification de la matrice d'adjacence se fait sous la forme d'un tableau de booléens qui permet de spécifier les connexions entre les nœuds. Par exemple pour la matrice de deux nœuds suivante (cf. tableau A.1), une connexion est créée entre le nœud source 'i' et le nœud destination 'j'.

	i	j
i	0	1
j	0	0

TABLE A.1: Exemple de matrice d'adjacence pour deux nœuds. Une connexion est créée entre le nœud source 'i' et le nœud destination 'j'.

Annexe B

Niveaux de spécifications de systèmes

Les niveaux de spécifications définits dans le chapitre 5 de l'ouvrage [ZKP00] sont présentés dans le tableau B.1. Nous présentons dans cette annexe chacun de ces niveaux de spécification en nous inspirant grandement du chapitre 5 de [ZKP00]. Nous considérerons le système « hôte d'une maladie contagieuse » pour illustrer l'un après l'autre ces différents niveaux de spécification. La figure B.1 présente la dynamique de l'infection et de l'excrétion et de la réaction immunitaire d'un hôte infecté par un agent pathogène [TDS⁺08].

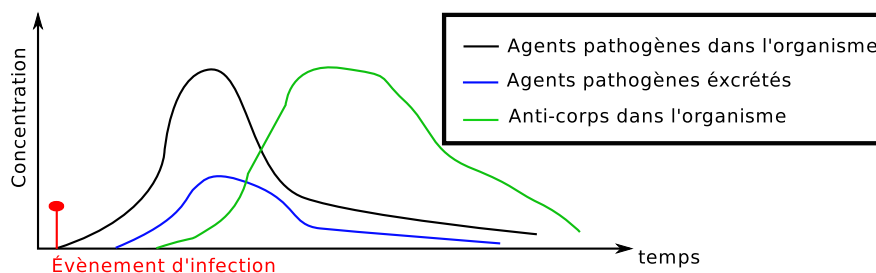


FIGURE B.1: Dynamique de multiplication et d'excrétion d'un agent pathogène infectieux par un hôte (superposition de graphiques de [TDS⁺08])

B.1 Cadre d'observation d'entrée/sortie (niveau 0)

Le niveau 0 est le cadre d'observation d'entrée/sortie désigné par les lettres IO pour Input et Output en anglais. À ce niveau un système est défini uniquement par la nature de ses entrées et de ses sorties, c'est à dire leurs domaines de définition. À ces deux ensembles, notés respectivement X et Y , il faut ajouter une base de temps T qui est un ensemble muni d'une relation d'ordre transitive, irréflexive et antisymétrique permettant de « dater » une entrée ou une sortie. Au niveau 0, notre système est donc spécifié formellement sous la forme de la structure suivante :

$$I/O = (T, X, Y)$$

où

T est une base de temps,

X est un ensemble - l'ensemble des valeurs d'entrée,

Y est un ensemble - l'ensemble des valeurs de sortie.

Dans le cas de notre exemple, la base de temps est l'ensemble des réels $T = \mathbb{R}$, le domaine de définition des entrées du système « hôte » est $X = \{0, 1\}$: à un instant donné, soit un évènement d'infection a lieu (entrée 1), soit il n'a pas lieu (entrée 0). Le domaine de définition des sorties du système « hôte » est $Y = \mathbb{R}$ l'ensemble des réels qui correspond au taux d'agents pathogènes excrétés à une date T . La figure B.2 schématise le système « hôte » spécifié au niveau IO.

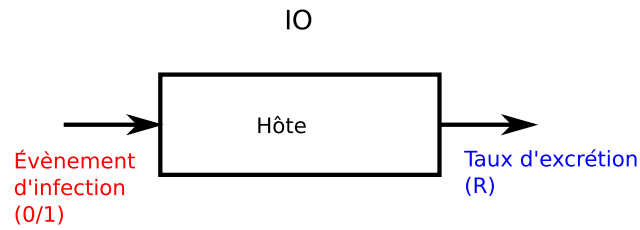


FIGURE B.2: Système « Hôte d'une maladie contagieuse » : Cadre d'observation d'entrée sortie.

L'entrée est binaire (à une date donnée soit il y a un évènement d'infection, soit il y en a pas). La sortie est continue (à une date donnée l'hôte excrète des agents pathogènes à un certain taux).

No	Spécification	Signification	Objet formel
0	Entrée/Sortie	Domaines des entrées/sorties	(T, X, Y)
1	Relation d'entrée/sortie	Ensemble des paires entrée/sortie	(T, X, Ω, Y, R)
2	Fonction d'entrée/sortie	Fonction qui à une entrée associe une seule sortie	(T, X, Ω, Y, F)
3	Système d'entrée/sortie	Description des transitions d'états	$(T, X, \Omega, Y, Q, \Delta, \Lambda)$
4	Spécification itérative	Description de fonctions permettant de générer la trajectoire de sortie et d'état en fonction de toute trajectoire d'entrée	$\langle T, X, \Omega_G, Y, Q, \delta, \lambda \rangle$
5	Système structuré	Une structure des ensembles et des fonctions permet d'identifier des composants	$(T, X, \Omega, Y, Q, \Delta, \Lambda)$ Où X, Y, Q, Δ et Λ sont structurés
6	Système multicomposants non modulaires	Système composé de plusieurs composants	$\langle T, X, \Omega, Y, D, \{M_d\} \rangle$
7	Réseaux modulaires (Systèmes couplés)	réseaux de systèmes spécifiés aux niveau 2 à 4	$\langle T, X_N, Y_N, D, \{M_d d \in D\}, EIC, EOC, IC \rangle$

TABLE B.1: Niveaux de spécification des systèmes dynamiques ([ZKP00])

B.2 Relation d'entrée/sortie (niveau 1)

Le niveau 1 est le comportement d'entrée/sortie, désigné par les lettres IORO pour (Input Output Relation Observation). On décrit à ce niveau la relation observée entre les entrées et les sorties d'un système. Nous attirons l'attention du lecteur sur le fait que la valeur de la sortie d'un système, observée à un instant donné, ne dépend pas nécessairement uniquement de la valeur de l'entrée du système mais peut dépendre aussi des valeurs d'entrées qui ont précédé. Ainsi, si l'on veut pouvoir décrire l'ensemble des relations d'entrées sorties observées sur un système, il faut mettre en relation des historiques (appelées aussi trajectoires) d'entrées avec des historiques (ou trajectoires) de sorties. Il nous est nécessaire à ce point d'introduire les notions de trajectoires et de segments.

Soit un ensemble A et une base de temps T , une trajectoire est une fonction $f : T \rightarrow A$. Il s'agit d'une fonction qui associe à chaque date, une valeur dans A . Un segment est simplement une trajectoire dont l'ensemble de définition est restreint à un intervalle de T . On note un segment défini entre les dates $t_1 \in \{\mathbb{R}, -\infty\}$ et $t_2 \in \{\mathbb{R}, +\infty\}$ de la manière suivante : $\omega :]t_1, t_2[\rightarrow A$, ou simplement $\omega_{<t_1, t_2>}$. La relation d'entrée/sortie que l'on va pouvoir observer est en fait la relation entre les segments d'entrées et les segments de sorties. Si l'on note respectivement ω_I et ω_O les segments d'entrée et de sortie, la relation R doit spécifier l'ensemble des couples $\{\omega_I, \omega_O\}$ que l'on peut observer. Considérons par exemple le cas de la relation d'entrée/sortie que l'on pourrait imaginer pour le système « hôte ». Il s'agirait de mettre en relation toute historique d'évènement d'infection possible avec les trajectoires d'excrétion qu'elles pourraient entraîner. La figure B.3 présente trois exemples de couples $\{\omega_I, \omega_O\}$ pour un système hôte imaginaire.

Formellement, un système se spécifie au niveau 1 sous la forme d'une structure :

$$IORO = (T, X, \Omega, Y, R)$$

Où T , X et Y ont la même signification que pour IO,

Ω est un ensemble - l'ensemble des segments d'entrées acceptables,

R est une relation - la relation d'entrée/sortie.

Ajoutons un mot sur l'ensemble Ω qui peut sembler superflu au premier abord, étant donné que l'ensemble X définit déjà les entrées qui sont acceptables en donnant l'ensemble de définition des variables d'entrées. L'ensemble Ω va simplement ajouter des précisions sur le type des segments d'entrées qui peuvent arriver au système. Pour donner un exemple, on peut imposer que les segments d'entrées soient des fonctions continues de temps. Imaginons que l'on considère un nouveau système hôte qui aurait une entrée « température » par exemple. Le domaine de définition des températures viables sera déjà spécifié au niveau 0 par l'ensemble X . Par contre au niveau 1, il est important de spécifier l'ensemble Ω sous la forme de l'ensemble des segments continus ou continus par morceaux¹ à valeur sur X puisque l'évolution de la température est une fonction continue du temps.

1. Un segment $\omega_{<t_1, t_2>} :]t_1, t_2[\rightarrow A$ est continu (respectivement continu par morceau) si la fonction $f : t \mapsto \omega_{<t_1, t_2>}(t)$ est continue, respectivement continue par morceau sur l'intervalle $]t_1, t_2[$.

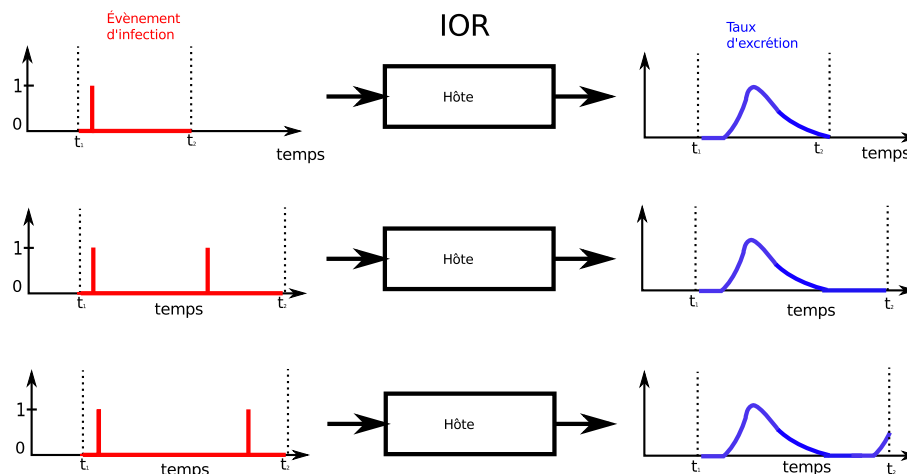


FIGURE B.3: Observations de relation d'entrée/sortie pour le système hôte

La figure présente trois expériences d'observations de la relation entrée/sortie du système hôte disposées sur trois lignes. À gauche en rouge les segments d'entrée et à droite en bleu les segments de sortie qui leur correspondent. Dans la première expérience l'hôte a été infecté une fois. Après une période d'incubation, il a excrété le pathogène pendant toute la période infectieuse puis a cessé d'excréter. Dans la seconde expérience il a été infecté deux fois mais la deuxième infection a eu lieu alors qu'il était encore immunisé et il n'a donc excrété le pathogène seulement pendant une période infectieuse. Dans la troisième expérience, l'hôte est également infecté deux fois, mais à un intervalle de temps plus grand. Cette fois, une nouvelle période infectieuse débute suite à la deuxième infection.

B.3 Fonction d'entrée/sortie (niveau 2)

Au niveau de spécification 1, on observe la relation d'entrée/sortie du système, c'est à dire que si l'on veut spécifier entièrement un système au niveau 1, on doit être capable de donner la liste des couples d'entrée/sortie $\{\omega_I, \omega_O\}$ observables sur le système. Cette spécification nous renseigne sur le système mais le lecteur observera qu'un segment d'entrée peut être associé à plusieurs segments sorties. La sortie du système dépend de l'entrée mais également de l'état interne du système que nous ne spécifions pas au niveau 1. Reprenons l'exemple du système hôte. Considérons les deux premières expériences présentées dans la figure B.3, c'est à dire les couples $\{\omega_I, \omega_O\}$ de la première ligne et de la deuxième ligne. Si l'on restreint la seconde expérience à sa deuxième moitié, on obtient un nouveau couple d'entrée/sortie $\{\omega_I < t'_1, t'_2 >, \omega_O < t'_1, t'_2 >\}$ pour la relation R. Ce nouveau couple est composé du même segment d'entrée que celui de l'expérience 1 mais d'un segment de sortie différent. Cela est dû à l'état interne du système hôte qui peut-être immunisé contre la maladie ou ne pas l'être. La figure B.4 présente ces deux expériences.

Au niveau de spécification 2, on doit être en mesure d'exprimer la relation d'entrée/sortie sous la forme d'une fonction, c'est à dire qu'un segment d'entrée doit pouvoir être associé à un seul et unique segment de sortie. Pour palier au fait que des sorties différentes peuvent être associées à un même segment d'entrée, on va

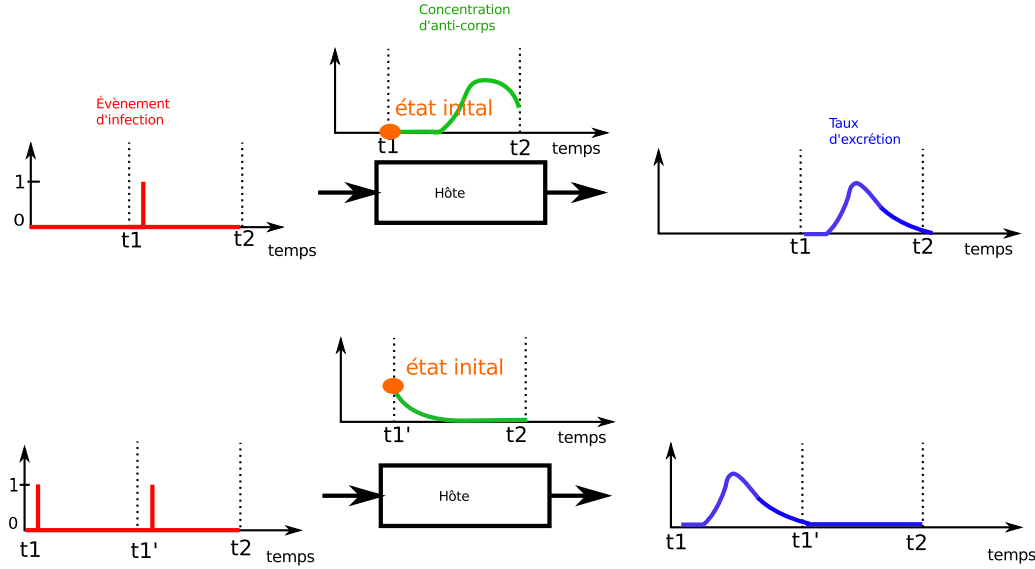


FIGURE B.4: Observation de la fonction d'entrée/sortie du système hôte. Pour pouvoir spécifier la fonction d'entrée/sortie d'un système (i.e. n'associer qu'un seul segment de sortie à un segment d'entrée) il faut prendre en compte un « état initial du système ». Dans le cas du système hôte on peut considérer la concentration d'anti-corps dans l'organisme de l'hôte. Les expériences présentées sur les ligne 1 et 2 présentent deux réponses différentes à une même entrée en fonction de l'état du système hôte (courbe verte représentant l'évolution de la concentration en anti-corps au cours de l'expérience). L'état initial est la valeur de la concentration en anticorps au début de l'expérience.

en fait définir un ensemble F de fonctions d'entrée/sortie. On considère que chaque fonction d'entrée/sortie peut être associée à un état initial dans lequel serait le système au début de l'expérience et qui serait déterminant sur la réponse du système. Dans le cas du système hôte, l'état initial serait « la concentration en anti-corps ». Un système est donc défini au niveau 2 par la structure suivante :

$$IOFO = (T, X, \Omega, Y, F)$$

Où T , X et Y et F ont la même signification que pour IOR, et F est un ensemble de fonctions d'entrées sorties avec la contrainte que

$$f \in F \Rightarrow f \subseteq \Omega \times (Y, T) \text{ est une fonction telle que si on note } \rho = f(\omega), \\ \text{dom}(\rho) = \text{dom}(\omega).$$

B.4 Système d'entrée/sortie (niveau 3)

Le concept d'état initial nous a permis de spécifier un système au niveau de la fonction d'entrée/sortie. C'est à dire qu'au niveau 2 on peut définir une fonction entre segments d'entrées et segment de sortie à condition de connaître l'état initial. Au niveau 3, on aimerait disposer d'une définition qui puisse traduire le fait que l'on puisse observer une succession d'expériences sur le système. C'est à dire qu'un système qui a vécu une expérience E va éventuellement répondre différemment que

le même système ayant vécu une expérience E' . Il est possible de prendre en compte cette succession d'expériences en spécifiant le nouvel « état initial » dans lequel le système est amené après avoir été perturbé par un segment d'entrée donné alors qu'il était dans un état initial donné. Reprenons l'exemple du système « hôte » et considérons les segments d'entrée/sortie $\langle t_1, t_2 \rangle$ qui apparaissent sur la deuxième ligne et que l'on note respectivement $\omega_I \langle t_1, t_2 \rangle$ et $\omega_O \langle t_1, t_2 \rangle$. Il s'agit de la concaténation des segments $\langle t_1, t'_1 \rangle$ et $\langle t'_1, t_2 \rangle$. Notons respectivement e_1 et e_2 les états initiaux des expériences 1 et 2. Ils sont marqués de points oranges sur la figure B.4. Imaginons maintenant que l'on enregistre les états finaux des deux expériences que l'on note respectivement f_1 et f_2 (c'est à dire la valeur de la concentration en anti-corps au temps t_2). On remarque que $e_2 = f_1$. En utilisant le système hôte spécifié au niveau 2 et la valeur de f_1 que l'on a enregistré, on peut donc dire que le segment d'entrée $\omega_I \langle t_1, t_2 \rangle$ et l'état initial e_1 vont entraîner la sortie $\omega_O \langle t_1, t_2 \rangle$ et amener le système dans l'état f_2 . On avance donc dans la définition du système hôte puisque l'on décrit des successions d'expériences. Formellement, le système d'entrée/sortie se définit par la structure suivante :

$$S = (T, X, \Omega, Y, Q, \Delta, \Lambda)$$

où T , X , Y et Ω ont la même signification que pour IOF,

Q est l'ensemble des états possibles

$\Delta : Q \times \Omega \rightarrow Q$ est la fonction de transition d'états globale,

$\Lambda : Q \times X \rightarrow Y$ (ou $\Lambda : Q \rightarrow Y$) est la fonction de sortie,

Avec certaines contraintes sur Ω et Λ qui assurent que pour tout segment d'entrée $\omega \in \Omega$, bien que les fonctions Λ et Δ renvoient des valeurs et non des segments, il soit possible d'arrêter et de relancer le système à toute date $t \in \text{dom}(\omega)$ de manière à connaître la valeur de Y et de Q à cette date, sans pour autant changer la valeur de sortie finale et l'état final obtenus pour la date $t_f = \max(\text{dom}(\omega))$.

B.5 Spécification itérative de système (niveau 4)

La spécification de niveau 4 permet de décrire toutes les expériences d'entrée/-sortie possibles que l'on peut réaliser sur un système, tout en décrivant également l'évolution de son état lors de ces expériences. Cependant elle n'est pas aisément utilisable en pratique. La fonction de changement d'état Δ par exemple doit être définie sur l'ensemble de tous les segments d'entrée possibles, ce qui la rend difficile à formuler en pratique. Pour cette raison, la spécification de système que l'on utilise en pratique utilise l'ensemble restreint de segments d'entrée Ω_G qui doit permettre de générer l'ensemble Ω par concaténations d'éléments de Ω_G . Nous avons déjà illustré le caractère itératif que permettait la spécification au niveau 4 dans le paragraphe précédent en prenant l'exemple du système hôte. Je passerai ici sur la démonstration de l'existence de Ω_G et sur sa définition en renvoyant le lecteur au chapitre 5 de [ZKP00] que nous avons déjà beaucoup repris dans cette section. La spécification itérative d'un système se fait donc sous la forme d'une structure :

$$G = \langle T, X, \Omega_G, Y, Q, \delta, \lambda \rangle$$

où T , X , Y et Q ont la même signification que pour S ,

Ω_G est l'ensemble des segments générateurs de Ω

$\delta : Q \times \Omega_G \rightarrow Q$ est la fonction de transition d'états pour des segments unitaires,

$\lambda : Q \times X \rightarrow Y$ (ou $\lambda : Q \rightarrow Y$) est la fonction de sortie.

C'est ce niveau de spécification que nous utiliserons pour spécifier des systèmes dynamiques non composés. Les trois niveaux de spécification qui suivent servent à décrire des systèmes composés, c'est à dire des systèmes dont on va spécifier d'une manière ou d'une autre la structure ou le comportement interne. Nous passerons moins de temps sur les systèmes composés qui ne sont pas toujours spécifiés formellement dans la thèse.

B.6 Système structuré (niveau 5)

On parle de système structuré lorsqu'il est possible de structurer les ensembles d'entrée/sortie, l'ensemble d'état et les fonctions de transition et de sortie. Il est spécifié de la même manière qu'un système S , sauf que X, Y, Q, Δ et Λ sont structurés.

B.7 Système multicomposants (niveau 6)

Un système multicomposants est une structure

$$MC = \langle T, X, \Omega, Y, D, \{M_d\} \rangle$$

où T, X, Y et Ω ont la même signification que pour S , D est l'ensemble des identifiants de composants, et pour tout $d \in D$,

$$M_d = \langle Q_d, E_d, I_d, \Delta_d, \Lambda_d \rangle$$

est un composant tel que

Q_d est l'ensemble des états du composant d ,

$I_d \subseteq D$ est l'ensemble des composants qui influencent d ,

$E_d \subseteq D$ est l'ensemble des composants influencés par d ,

$\Delta_d : \times_{i \in I_d} Q_i \times \Omega \rightarrow \times_{i \in E_d} Q_i$ est la fonction de transition d'état de d ,

$\Lambda_d : \times_{i \in I_d} Q_i \times \Omega \rightarrow Y$ est la fonction de sortie de d .

Nous 'avons pas eu besoin de spécifier formellement de systèmes au niveaux 5 et 6 pour défendre la thèse énoncée dans notre problématique. Nous verrons que le caractère modulaire du niveau 7 nous ont amené à privilégier ce niveau pour la spécification de systèmes structurés. Cependant, il nous semble important de les évoquer ici. Les modèles individus-centrés formalisés sous forme mathématique sont souvent spécifiés à ces niveaux de spécification. On peut découper l'état du système en une partition dont chaque sous ensemble correspondrait à l'état d'un individu. Un ensemble d'équations différentielles permet alors de décrire l'évolution de l'état du système (correspondant à la fonction Δ) [TDZ08]. Suivant la manière dont on spécifie cette fonction de transition, on va se retrouver au niveau 5 ou au niveau 6.

Ces niveaux de spécification sont intéressants à nos yeux car ils permettent d'utiliser des méthodes mathématiques pour raisonner sur des comportements du système en liaison avec les composants de ce système qui ont été identifiés. Je citerai deux

exemples. Le premier est un exemple dans lequel un algorithme d'apprentissage par renforcement est utilisé à l'échelle d'un système composé, tout en prenant en compte des récompenses définies au niveau des composants [CB05]. Dans ce cas, le système est spécifié au niveau 5 et le fait de disposer de la fonction de transition d'états de l'ensemble du système facilite grandement les calculs. Le second exemple est un travail qui consistait à utiliser des méthodes d'approximation par moments pour simplifier un modèle individu-centré de propagation d'une maladie contagieuse dans une population d'individus en mouvement (cf. annexe E). Le système était spécifié au niveau 6 et la formulation des fonctions de transition de l'état d'un composant en fonction de l'état des autres nous a permis d'utiliser les méthodes d'approximation par moments.

B.8 Systèmes couplés structurés (niveau 7)

Au niveau de spécification de systèmes couplés (on parle aussi de réseau de systèmes), chaque composant peut être spécifié lui même sous la forme d'un système d'entrée/sortie (S) et on définit un ensemble de relations entre les entrées et les sorties du système global et des systèmes composants ainsi qu'entre les entrées et sorties des systèmes composants entre eux.

Un réseau de systèmes structuré se définit de la manière suivante :

$$N = \langle T, X_N, Y_N, D, \{M_d | d \in D\}, EIC, EOC, IC \rangle$$

où

X_N est l'ensemble des ports d'entrées $IPorts_N$ de N et des ensembles de définition des variables d'entrées correspondantes (chaque port est associé à une variable et son ensemble de définition),

Y_N est l'ensemble des ports de sortie $OPorts_N$ de N et des ensembles de définition des variables de sortie correspondantes (chaque port est associé à une variable et son ensemble de définition),

D est l'ensemble des identifiants des composants,

$M_d = \langle T, X_d, Y_d, Q_d, \Omega_d, \Lambda_d \rangle$, est un système tel que X_d et Y_d sont les ensembles de ports d'entrées et de sorties $IPorts_d$ et $OPorts_d$ associés aux ensembles de définitions des entrées et des sorties du composant d,

$EIC \subseteq \{((N, ip_N), (d, ip_d)) | ip_N \in IPorts_N, ip_d \in IPorts_d\}$ est le couplage au entrées externes,

$EOC \subseteq \{((N, op_N), (d, op_d)) | op_N \in OPorts_N, op_d \in OPorts_d\}$ est le couplage au sorties externes,

$IC \subseteq \{((a, op_a), (b, ip_b)) | (a, d) \in D^2, op_a \in OPorts_a, ip_b \in IPorts_b\}$ est le couplage interne.

Remarquons que la structure N est également un système. On peut donc tout à fait spécifier un des systèmes composants $\{M_d | d \in D\}$ sous la forme d'un réseau de systèmes lui-même. On obtient alors une structure emboîtée dite hiérarchique telle qu'illustrée figure B.5.

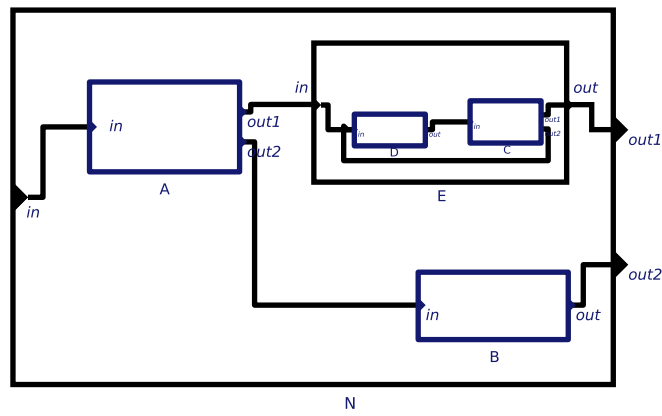


FIGURE B.5: Hiérarchie de systèmes couplés.

N est un exemple de système couplé dont un des composants (le sous-système E) est lui même un système couplé.

Annexe C

Modèles composants de T^*

C.1 Modèle d'observation passive retardée

La figure C.1 montre comment le modèle d'observation passive peut être couplé à un modèle de délai pour créer un modèle composé d'observation passive avec délai.

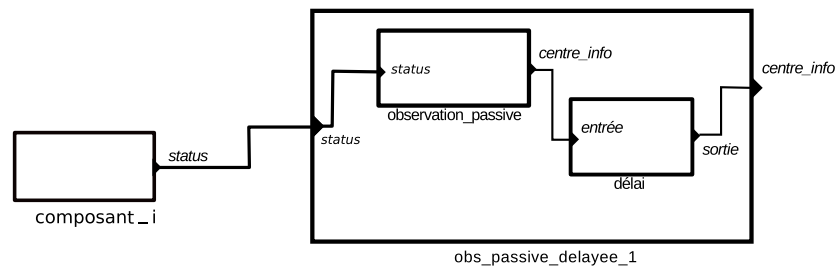


FIGURE C.1: Schéma de connexion entre un modèle d'observation delayée et un modèle composant du système cible.

C.2 Modèle d'observation planifiée

La figure C.2 montre comment le modèle d'observation réactive peut être couplé à un modèle ordonnanceur pour créer un modèle composé d'observation planifiée.

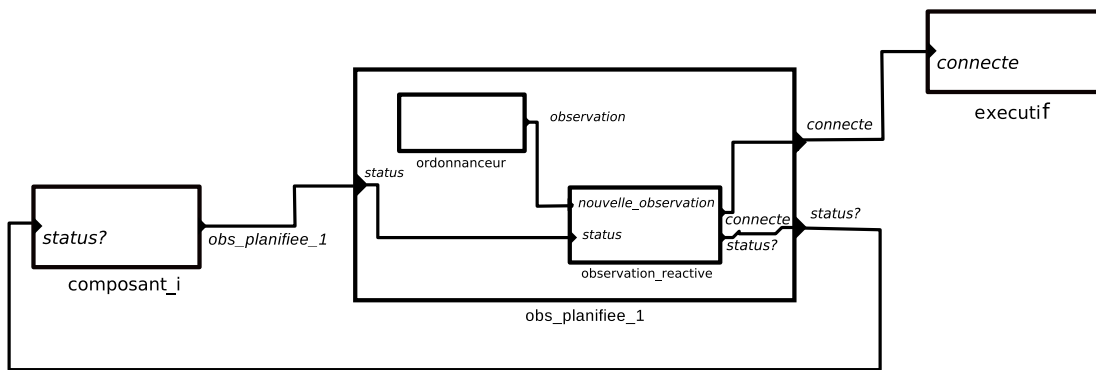


FIGURE C.2: Schéma de connexion entre un modèle d'observation planifiée et un modèle composant du système cible.

Annexe D

MPE d'estimation de paramètres utilisant une méthode d'optimisation par essaim de particules

La méthode d'Optimisation par Essaims de Particules (OEP) (traduction de *Particle Swarm Optimisation (PSO)* en anglais) est une méta-heuristique qui consiste à explorer un espace de recherche en considérant un ensemble de vecteurs de cet espace comme un essaim de particules. Une position et une vitesse de départ sont attribuées à chaque particule que l'on fait se déplacer dans l'espace, tout en réévaluant régulièrement la vitesse de chaque particule en fonction de la meilleure position rencontrée par l'essaim et par la particule elle-même. L'algorithme sur lequel nous nous sommes basé a été proposé dans [KE95]. Nous avons utilisé la version donnée dans [Ped10] qui est rappelée dans l'encadré 1 et illustrée figure D.1.

À titre d'exemple de MPE utile dans la triade de Minsky, nous avons développé un MPE qui utilise la méthode d'Optimisation par Essaims de Particules (OEP) pour estimer par simulation les paramètres d'un modèle. Notons ce modèle MPE_{OEP} . Le MPE_{OEP} est générique. C'est à dire qu'il est possible de le paramétrer pour l'estimation de paramètres continus (ou de valeurs initiales de variables d'états continues) de n'importe quel MSD qui soit formalisé dans un formalisme conforme à DEVS. C'est cette généricité qui nous intéresse ici car elle illustre la capacité de notre cadre à modéliser et simuler une triade de Minsky. La qualité des estimations fournies par le MPE_{OEP} dépend des cas d'utilisations (valeurs des paramètres $N, \omega, \phi_p, \phi_g$ de l'algorithme d'OEP, observations à fitter, surface de réponse du modèle, critère de différence aux données utilisés, etc.).

Minimisation de la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ par un essaim de N particules.

Notations :

$\vec{x}_i \in \mathbb{R}^n$ est la position de la particule i ,
 $\vec{p}_i \in \mathbb{R}^n$ est la meilleure position rencontrée par la particule i ,
 $\vec{v}_i \in \mathbb{R}^n$ est la vélocité de la particule i ,
 \vec{b}_{low} et \vec{b}_{up} sont les limites basse et haute de l'espace de recherche,
 $\vec{d} = |\vec{b}_{up} - \vec{b}_{low}|$,
 \vec{g} est la meilleure position rencontrée par l'essaim.
 ω , ϕ_p et ϕ_g sont des paramètres de l'algorithme.

Initialisation :

- 1: **for** $i = 1 \rightarrow N$ **do**
- 2: $\vec{x}_i \leftarrow U(\vec{b}_{low}, \vec{b}_{up})$
- 3: $\vec{p}_i \leftarrow \vec{x}_i$
- 4: $\vec{v}_i \leftarrow U(-\vec{d}, \vec{d})$
- 5: **end for**
- 6: $\vec{g} \leftarrow \vec{x}^*$ où \vec{x}^* est la position de x_i telle que $f(\vec{x}_i)$ est minimum.

Itération jusqu'à un critère d'arrêt :

- 1: **for** $i = 1 \rightarrow N$ **do**
- 2: $r_p \leftarrow U(0, 1)$
- 3: $r_g \leftarrow U(0, 1)$
- 4: $\vec{v}_i \leftarrow \omega \vec{v}_i + \phi_p r_p (\vec{p}_i - \vec{x}_i) + \phi_g r_g (\vec{g} - \vec{x}_i)$
- 5: **for** $j = 1 \rightarrow n$ **do**
- 6: $v_i^j \leftarrow \text{bound}(v_i^j, -d_i^j, d_i^j)$
- 7: **end for**
- 8: $\vec{x}_i \leftarrow \vec{x}_i + \vec{v}_i$
- 9: **for** $j = 1 \rightarrow n$ **do**
- 10: $x_i^j \leftarrow \text{bound}(x_i^j, b_{low_i}^j, b_{up_i}^j)$
- 11: **end for**
- 12: **end for**
- 13: **for** $i = 1 \rightarrow N$ **do**
- 14: **if** $f(\vec{x}_i) \leq f(\vec{p}_i)$ **then**
- 15: $\vec{p}_i \leftarrow \vec{x}_i$
- 16: **end if**
- 17: **if** $f(\vec{x}_i) \leq f(\vec{g})$ **then**
- 18: $\vec{g} \leftarrow \vec{x}_i$
- 19: **end if**
- 20: **end for**

Encadré 1 : Optimisation par Essaim de Particules

Le MPE_{OEP} est un modèle DEVS couplé composé de trois modèles DEVS atomiques : un modèle de générateur de plan d'expériences spécifique (noté G_{OEP}), le modèle expérimenteur générique tel qu'il est spécifié dans la section précédente, et un modèle d'analyse de résultat générique (noté A_{OEP}). Ces trois modèles sont couplés selon la forme générique des MPE présentée figure 3.6. Le modèle doit recevoir sur son port « input » en entrée des événements d'observations (notés $y \in \mathbb{R}$) et

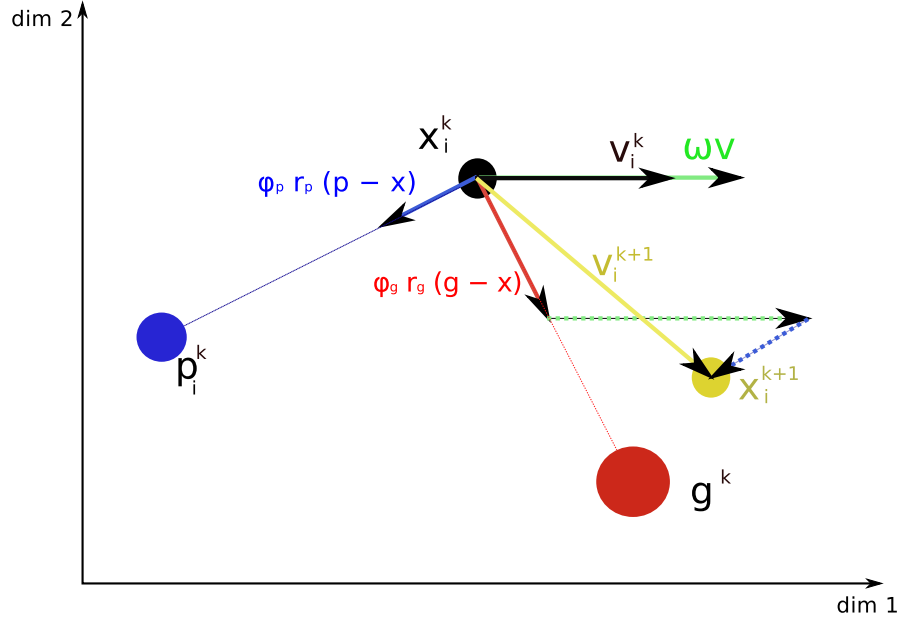


FIGURE D.1: Déplacement d'une particule dans un Essaim de Particules pour un problème en deux dimensions. À une itération « k », une particule « i » a une position x_i^k une vitesse v_i^k et une « meilleure position rencontrée » p_i^k . La meilleure position rencontrée par l'essaim est g^k . La nouvelle vitesse de la particule est calculée par la formule : $v_i^{k+1} = \omega v_i^k + \phi_p r_p (p_i^k - x_i^k) + \phi_g r_g (g^k - x_i^k)$. Sa nouvelle position est calculée par la formule : $x_i^{k+1} = x_i^k + v_i^{k+1}$.

renvoyer sur son port « output » la liste des valeurs estimées pour chaque paramètre (notée $\theta^* \in \mathbb{R}^n$ si on a n paramètres à estimer). De manière plus formelle, on peut donner la spécification du cadre d'entrée/sortie d'un modèle MPE_{OEP} pour une estimation de « n » paramètres :

$$IO_{MPEOEP} = \langle X_{MPEOEP}, Y_{MPEOEP} \rangle$$

où

$$\begin{aligned} X_{MPEOEP} &= \{\{input\}, \{\mathbb{R}\}\}, \\ Y_{MPEOEP} &= \{\{output\}, \{\mathbb{R}^n\}\}. \end{aligned}$$

Décrivons succinctement les points clefs du fonctionnement du MPE_{OEP} . Tout d'abord, rappelons que la définition du modèle à calibrer (le modèle dont on veut estimer les paramètres) est contenue dans l'état du MPE_{OEP} . Elle fait plus précisément partie de l'état du modèle G_{OEP} qui l'utilise pour construire la structure de plan d'expériences PE envoyée au modèle expérimenteur (cf. section 3.3.3). Le modèle à calibrer (notons le M_C) doit répondre à certains critères : il doit s'accommoder au cadre expérimental CE_C que nous décrivons dans la figure D.2. Le cadre expérimental CE_C est lui-même un modèle DEVS couplé. Il est composé de deux modèles atomiques. Le premier est un générateur d'évènements qui dans notre cas se limite à perturber le modèle M_C à l'initialisation pour le mettre dans l'état initial voulu. Le second est un accepteur d'évènements qui possède un paramètre nommé « dates d'observations » et qui va en fait interroger le modèle M_C sur son état à ces

dates et récupérer et stocker l'état du modèle M_C à ces dates. Le modèle défini dans les expériences d'une structure PE générée par G_{OEP} est le modèle couplé composé d'un modèle M_C et d'un cadre expérimental CE_C . On appelle ce modèle couplé le « modèle cadré » en référence au terme de « *framed model* » utilisé dans [Tra06].

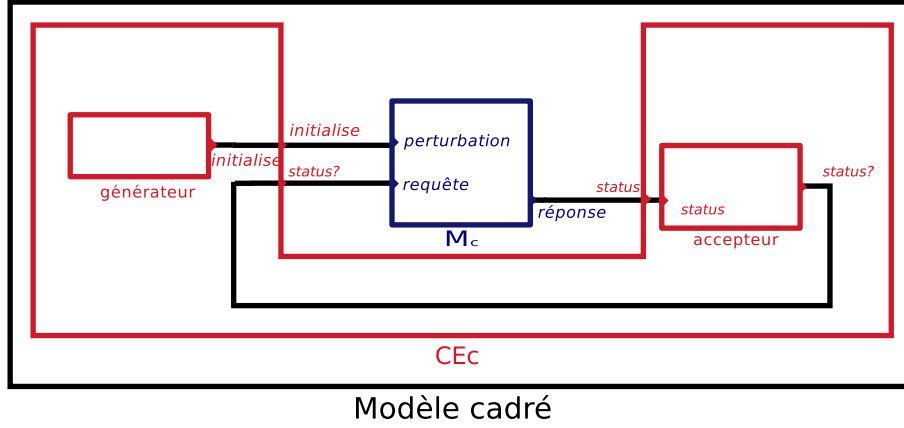


FIGURE D.2: Cadre expérimental pour l'estimation de paramètres. Le modèle à calibrer M_C doit s'accommoder du cadre expérimental CE_C . C'est à dire qu'il doit avoir un port « perturbation » permettant d'initialiser les variables d'états ou paramètres à estimer, et un port « requête » et un port « réponse » qui lui permettent de générer un évènement de sortie avec la valeur de son état (un nombre réel) si un évènement est reçu sur le port « requête ». On appelle « modèle cadré » le modèle couplé composé du M_C et du CE_C .

Le modèle G_{OEP} reçoit des observation $(y(t_l)) \in \mathbb{R}$ sur son port d'entrée *input*. À chaque nouvelle observation reçue, il la stocke et construit ainsi une série temporelle $(y(t_l))_{l=1..N}$. Une fois la nouvelle observation stockée, une *fonction de déclenchement* est appelée qui détermine si une phase de calibration est déclenchée. Dans le cas d'application présenté dans le chapitre suivant, par exemple, la phase de calibration est déclenchée à partir du moment où le nombre d'observations dans la série temporelle est supérieur à 5.

Lorsque la phase de calibration est déclenchée, le modèle G_{OEP} génère les particules en suivant la procédure d'initialisation de l'algorithme OEP décrite dans l'encadré 1 jusqu'à la ligne 5. À ce point une première phase d'évaluation a lieu. Elle se déroule de la manière suivante.

Le modèle G_{OEP} construit un plan d'expériences PE composé d'autant d'expériences qu'il y a de particules. Dans chaque expérience, il affecte les valeurs d'état initiaux ou de paramètres de M_C à estimer selon la position de la particule correspondante. Il affecte également la liste des dates auxquelles ont été reçues les observation $((t_l)_{l=1..N})$ au paramètre « dates d'observations » du modèle accepteur du CE_C . Le plan d'expérience est ensuite envoyé au modèle expérimenteur, qui génère une structure de résultats d'expériences RE et la transmet au modèle A_{OEP} qui la transmet à son tour au modèle G_{OEP} qui la reçoit sur son port *feedback*. Pour chaque particule « i » le modèle G_{OEP} dispose ainsi de la liste des valeur de l'état du modèle M_C simulé avec les paramètres x_i à chacune des dates $((t_l)_{l=1..N})$ observées par le modèle accepteur. Une des mesures de distance D du tableau 2.1 est alors

ANNEXE D. MPE D'ESTIMATION DE PARAMÈTRES UTILISANT UNE
MÉTHODE D'OPTIMISATION PAR ESSAIM DE PARTICULES

Fonctions	Arguments	Sortie
fonction de déclenchement	série temporelle	booléen
fonction de distance	deux série temporelle	réel
critère d'arrêt	-	booléen
Paramètres	Type	
ω	\mathbb{R}	
ϕ_g	\mathbb{R}	
ϕ_p	\mathbb{R}	
N	\mathbb{N}	
\vec{b}_{low}	\mathbb{R}^n	
\vec{b}_{up}	\mathbb{R}^n	

TABLE D.1: Fonctions et paramètres à définir pour utiliser un modèle MPE_{OEP} pour un problème à « n » dimensions.

utilisée pour évaluer la valeur de $f(\vec{x}_i)$ comme étant $D(\vec{x}_i, (y(t_l))_{l=1..N})$. La meilleure position de l'essaim peut alors être mise à jour (ligne 6 de la phase d'initialisation de l'algorithme EOP).

Ensuite la phase d'itération de l'algorithme OEP est itérée jusqu'à ce qu'un « critère d'arrêt » soit rencontré. Une itération se déroule en deux temps. D'abord les particules sont déplacées en exécutant les lignes 1 à 12 de la phase d'itération de l'algorithme OEP. Ensuite une phase d'évaluation a lieu. Enfin les meilleures positions des particules et de l'essaim peuvent être mises à jour en exécutant les lignes 13 à 20.

Le tableau D.1 résume l'ensemble des fonctions et des paramètres à définir pour utiliser un EPM_{OEP} .

Annexe E

Approximation par moments d'un modèle épidémiologique individu-centré

Moment approximation of the infection dynamic in a population of moving host

Bruno Bonté¹, Jean-Denis Mathias², Raphaël Duboz^{1,3}

¹CIRAD
UR22 Animal et
Gestion Intégrée des
Risques
Campus International de
Baillarguet
34398 Montpellier cedex 5,
France

²CEMAGREF
LISC
24 avenue des Landais,
B.P. 50085, 63172
Aubière,
France

³AIT
Asian Institute of
Technology
P.O. Box 4, Klong Luang
Pathumthani 12120
Thailand

Keywords : *Individual Based Model, Disease Spread, Mobility, Aggregation method, H5N1.*

Abstract

The modelling of contact networks is of key importance in the field of epidemiology. Currently, studies mainly focused on networks with stationary structures unless we know those structures being dynamic with a constant appearance and disappearance of links over time. It is particularly true for the contact network of moving individuals, where the displacement behaviours generate the structure of the contact network. Individual Based Models (IBM) can simulate the individual behaviours involved in the contact process. However, if we consider very large populations, IBM can be hard to simulate and to study due to the computational costs. In this paper, we use the Moment Approximation (MA) method to approximate a stochastic IBM with an aggregated and deterministic model. We illustrate the method considering an application in animal epidemiology, the spread of the highly pathogenic virus H5N1 of Avian Influenza in a poultry flock. The MA method is explained in a didactic way such as it can be reused and extended. We compare simulation results of three models. The IBM, the model that we obtain after the moment approximation (the first and second moments are simulated), and the model that we obtain after the mean field approximation (only the first moment is simulated). We call these models respectively, the IBM, the MA model, and the MF model. Results show a good agreement between the MA model and the IBM. They highlight the importance for the models to capture the displacement behaviors and the spatial structure in the study of disease spread. Furthermore, this paper illustrates an

original way of using different models of the same system to learn more about the system itself, and about the representation we build of it.

1 Introduction

The structure of the contact network of individuals plays an important role in the dynamic of spread of infectious diseases. It has been shown that different contact structures lead to different dynamics of infection (see [1, 2] for recent works on this subject). We know that mean field models, which make the hypothesis of a constant and homogeneous contact rate between susceptible individuals and infectious individuals [3], can not always be considered as a acceptable approximation. Indeed, clusters of infected individuals appear during the epidemic course implying that the infectious rate can exhibit complex spatio-temporal dynamics [4, 5], depending on the structure of the contact network [6]. For particular situations and scales, the structure of the contact network should be taken into account [7, 1, 8].

For these reasons, many recent models consider the spread of infectious diseases as a diffusion process within a contact network [9, 10], raising the issue of being able to draw the contact network. In some cases, the network is built from population surveys at the appropriate scale (see [11] for an example in human epidemiology and [12] in animal epidemiology). Nevertheless, the contact network is difficult to design. Very often, the entire population can not be sampled and the network building relies on descriptive statistics. Therefore, a current issue in the disease spread modelling field is to accurately reproduce the infection dynamics at the population level, having a minimum information about the structure of the contact network between individuals. To do that, we can use models that approximate the impact of the network structure on the spread of the disease [6, 13]. Unfortunately, structure statistics themselves are not always available. It is often the case in animal epidemiology where the protocols used in ecology to build these statistics are not always able to produce reliable statistics for epidemiology [14]. To tackle this issue, we can simulate the individual movements and then generate potential contact patterns between individuals. Such Individual Based Models (IBM) have been done for disease transmission [15, 16, 17]. IBM are mechanist and often stochastic models facing several difficulties regarding their rigorous communication [18] and their sensitivity analysis and parameters estimation [19]. Furthermore, IBM can not be analytically analysed at the population scale to study asymptotic behaviours.

In order to find a trade-off between simulated contact patterns using an IBM and the absence of reliable network data, we propose to approximate the IBM with a deterministic and mathematically tractable model using the moment approximation method [20, 21]. This method has already been used to approximate models of disease spread in networks [8, 22, 23] where the structure of contacts is known. In this paper, we approximate an IBM that explicitly simulates the individual movements. Doing that, we take into account the dynamic of contact structures. The aim of this paper is to test if such an approximation

is valid and to discuss the impact of individual movements on epidemic dynamics. To do that, we compare simulation results of three models. The IBM, and two approximations of the IBM. The first approximation is the moment approximation, we simulate up to the second moment of the IBM dynamics and we approximate the third moment. We will call the resulting model, the MA model (or just MA), in order to facilitate the reading. The second approximation is the mean field approximation which is usually used in epidemiology. We explain later in the paper that this approximation corresponds to simulate the first moment of the IBM dynamics and to approximate the second moments using the first moment. Once again, in order to facilitate the reading, we call this model the MF model (or just MF).

In this paper, we choose the example of the Highly Pathogenic virus H5N1 spreading in a poultry flock. H5N1 virus is of great interest for civil society due to the risk of zoonosis. We think that modelling an epidemic spread in a poultry flock is an interesting example because it gives us an epidemic system with moving individuals which can be simulated with a limited number of rules.

In the first part, we describe the IBM focusing on the mathematical formulation of processes at the individual level. In a second part, we build the moment approximation model (MA). In a third part, we recall that the mean field epidemiological model (MF) is the first moment approximation model. In a fourth part, we compare simulation results of the three models: IBM, MA and MF.

2 Individual Based Model

We consider a population of N individuals. Each individual is identified by a unique number i with $1 \leq i \leq N$ and has two attributes: its position in a two dimension space ($x \in \mathbb{R}^2$) and an epidemiological state (noted $s \in \{S, E, I, R\}$). Epidemiological state divides the host population in four sub-populations of individuals: susceptible individuals (S) which are free of disease but can get infected when exposed to the virus, exposed individuals (E) which have been infected but are not infectious yet (they are multiplying the virus but do not shed it), infectious individuals (I) that are shedding the virus and can transmit the infection to susceptible individuals, and removed individuals (R) which have been infected and are either dead or have recovered from infection and are permanently immunized. We will use formal notations to describe the IBM (they come from [24]). Position and epidemiological states of individual i at time t are respectively noted x_i^t and s_i^t .

Model dynamics includes 4 processes: individuals move, infection of susceptible individuals by infectious individuals, incubation process for exposed individuals that tells when they become infectious and start shedding the virus and recovery or death of infectious individuals that make them removed individuals.

2.1 Individual move

Susceptible, exposed and infectious individuals move the same way. We assume that each individual moves following a kernel function ω_1 . The kernel function can be Gaussian or simply a local window. In our case, the kernel is a local uniform kernel. The instantaneous probability that, during an infinitesimal time dt , an individual has moved to any location x from its previous location x_i^t is given by the function P_M with:

$$\forall x, \forall i, P_M(x_i^{t+dt} = x) = \lambda \omega_1(x - x_i^t) \quad (1)$$

Note that $\lambda \in [0, 1]$ represents the probability that an individual moves.

2.2 Infection process

It is the process that makes a susceptible individual becoming exposed due to infectious contact with infectious individuals.

We assume that the expected contact rate between two individuals i and j is given by a kernel function $\omega_2(x_i - x_j)$. We hence consider that at any time a contact structure exists (at any time any individual i is linked to any individual j by a valued link with the value $\omega_2(x_i - x_j)$). We also assume that each contact with an infectious individual has a probability "c" of getting him exposed to the virus. Let's write $P_{E\Delta t}(s_i^{t+\Delta t} = E | s_i^t = S)$ the probability that the susceptible individual i get infected during Δt .

Then we can compute k_I the expected contact rate with infectious individuals as the following:

$$k_I = \sum_{j|s_j=I} \omega_2(x_i - x_j)$$

And the expected number of contact with infectious individuals during Δt is $k_I \times \Delta t$.

Then, the probability that a susceptible individual i is not infected during Δt is the following ([26]):

$$1 - P_{E\Delta t}(s_i = E) = (1 - c)^{k_I \times \Delta t} \quad (2)$$

Actually, what we want is the instantaneous probability of infection namely $P_E(s_i^{t+dt} = E | s_i^t = S)$. So we compute the limit of equation 2 when Δt goes to zero, and we get the following expression of P_E (equation 3).

$$P_E(s_i^{t+dt} = E | s_i^t = S) = \beta \sum_{j|s_j^t=I} \omega_2(x_i^t - x_j^t) \quad (3)$$

where $\beta = -\log(1 - c)$. This limit computation, as well as the whole hypothesis and principles leading to equation 3 are quite usual in epidemiology (details can be found in [26] chapter 2 box 2.1.). We gave their details here because we want to describe how the contact structure is computed through the distances between individuals, and how it drives the infection process. We will see in section 4 how this structure disappears in the mean field model.

2.3 Incubation process

The period during which an individual is exposed but not infectious yet, is called the latency period. Incubation process (virus multiplication inside a host) is summarized by the fact that at the end of the latency period, an exposed individual i ($s_i = E$) becomes infectious ($s_i = I$). Hence, we only represent the event leading to the change of the individual epidemiological state. The instantaneous probability that this event occurs, noted P_I , is constant over the latency period as shown by equation 4.

$$P_I(s_i^{t+dt} = I | s_i^t = E) = \alpha \quad (4)$$

where α is a disease latency coefficient expressed in $time^{-1}.individuals^{-1}$.

2.4 Removing process

Recovery or death of infectious individuals leads to the removal of the recovered or dead individuals. Similarly as for the incubation process, instantaneous probability P_R that an infectious individual gets removed is the following:

$$P_R(s_i^{t+dt} = R | s_i^t = I) = \gamma \quad (5)$$

where γ is the recovery rate coefficient expressed in $time^{-1}.individuals^{-1}$.

Model discretization used for implementation is presented in appendix section A.

3 Moment approximation

From the above processes, we investigate now how the contact pattern between individuals changes. We suppose that this pattern is stochastic and Markovian. The idea is to compute the mean dynamics of the susceptible, exposed, infectious and recovered individual densities and the mean dynamics of the correlations between individuals. For this purpose, a moment approximation is used. This approach has been first employed in fluid mechanics [27]. Then, it has been used in order to simplify ecology models ([20]). The approach used in this study is inspired by [20] and [21]. We need to discuss about two hypothesis required by this method. First, we suppose that we have a toroidal space. It leads to suppose that we simulate a representative area of our space. Second, this kind of method suppose that we have a large number of individuals in our dynamics. Indeed, we calculate the mean dynamics and it is more representative with a lot of individuals in order to avoid to perform a large number of simulations. The mathematical details are available in the references cited above. This method is employed here in order to highlight the influence of the spatial structure on the disease spread. The dynamics of mean population densities are first developed.

3.1 Dynamics of mean population densities

As a first step, we write down the dynamics of the mean population densities of susceptible, exposed and infectious individuals (first moments) that we note N_S , N_E and N_I . The equations of these dynamics are derived as in [20]. They include the contributions of the processes defined previously, which modify the mean densities. The evolution of N_S is defined by:

$$\frac{dN_S}{dt} = -\beta \int \omega_2(\xi) C_{SI}(\xi) d\xi \quad (6)$$

where $C_{SI}(\xi)$ is the average "susceptible individuals - infectious individuals" pair density with locations separated by ξ . For a given set of individuals and a given configuration ξ , the pair density is defined as:

$$c_{SI}(\xi) = \sum_{i=1}^{N_S} \sum_{j=1}^{N_I} \delta(x_i^S - x_j^I - \xi) \quad (7)$$

with $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ otherwise. $C_{SI}(\xi)$, is the average of the function $c_{SI}(\xi)$ over all possible realizations of susceptible and infectious distributions.

Similarly, for the exposed individuals, the mean dynamics writes:

$$\frac{dN_E}{dt} = \beta \int \omega_2(\xi) C_{SI}(\xi) d\xi - \alpha N_E \quad (8)$$

Then, for the infectious individuals, the mean dynamics writes:

$$\frac{dN_I}{dt} = \alpha N_E - \gamma N_I \quad (9)$$

We can see that the equations ruling the evolution of the mean number of susceptible, exposed and infectious individuals are very similar. The infection part is linked to local density of infectious individuals around the susceptible individuals. Now, the dynamics of equations 6 and 8 depends on the pattern of the correlation densities calculated as in equation 7 and on its dynamic which is calculated by the second moment (section 3.2). Note that we consider a static number of individuals N and hence, the number of removed individuals can be simply computed as follows:

$$N_R = N - (N_S + N_I + N_E) \quad (10)$$

As can be shown in the equations of the mean densities (see equations 6 and 8), the correlation densities play an important role in these dynamics. These density dynamics are now developed in the following section.

3.2 Dynamics of pair correlation densities

The second moment dynamics corresponds to the pair correlation density dynamics. Most important is to evaluate the dynamics of susceptible and infectious individual pair correlation densities (C_{SI} as defined by equation 7). Indeed, the dynamics of the mean densities of susceptible and exposed individuals depend on this dynamics (see equations 6 and 8). This dynamics allows us to consider indirectly the contact network. Next paragraph presents the equation and shows how we built it from the IBM mathematical specification. To do that, we needed first to calculate the dynamics of all pairs of individuals. As explained above, we focus here on the C_{SI} which is more informative. Other pair correlation dynamics, as well as triplet dynamics are described in appendices (see appendix B).

$$\begin{aligned} \frac{dC_{SI}(\xi, t)}{dt} = & \\ \text{Movement} \left\{ \begin{aligned} &+ 2\lambda \int \omega_1(\xi') C_{SI}(\xi + \xi') d\xi' \\ &- 2\lambda |\omega_1(\xi)| C_{SI}(\xi) \end{aligned} \right. & \\ \text{Infection} \left\{ \begin{aligned} &- \beta \int \omega_2(\xi') T_{SII}(\xi, \xi') d\xi' \\ &- \beta \omega_2(\xi) C_{SI}(\xi) \end{aligned} \right. & \\ \text{Recovery} \{ &- \gamma C_{SI}(\xi) \\ \text{Latency} \{ &\alpha C_{SE}(\xi) \end{aligned} \quad (11)$$

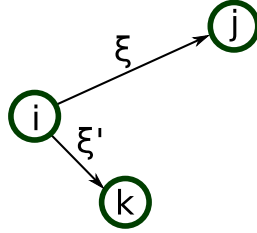


Figure 1: $T_{ijk}(\xi, \xi')$ triplet configuration. Individuals i and j are separated by a distance ξ , and individuals i and k by a distance ξ'

where $T_{ijk}(\xi, \xi')$ represents the correlation density of triplets as shown on Figure 1. The infectious-susceptible pair correlation dynamics depends on the four processes in the above equation:

- the movement terms can be decomposed in two terms. The first term computes the expected number of new pairs. A pair at distance $\xi + \xi'$ is here considered. When an individual of this pair moves at a distance ξ' , a

new pair is created with a distance ξ . The second computes the expected number of destroyed pairs. When an individual of an (SI) -pair at distance ξ moves, it destroys the pair. These terms are multiplied by 2 because both individuals have a chance to move;

- the infection terms correspond to the exposition to virus of the susceptible individual S of the (SI) -pair. There are two possibilities. The first possibility depends on the triplet configuration between a susceptible and two infectious individuals (T_{SII}) and on the interaction kernel ω_2 . When the susceptible individual of a (SI) pair is infected by an other infectious individual located at distance ξ' , it destroys the pair. The second possibility lies in a too short distance of the susceptible and the infectious individual of a (SI) -pair, following the interaction kernel ω_2 . In this case, the (SI) -pair is destroyed by the infection of the susceptible;
- the recovery term computes the number of infectious individuals in situation ξ with respect to a susceptible individual. This individual has a probability γ to recover and thus to destroy a susceptible-infectious pair;
- the latency term corresponds to the number of exposed individuals that are in a configuration distant of ξ with a susceptible individual and hence create a new (SI) -pair distant of ξ if they become infectious.

The description of the $C_{SI}(\xi, t)$ dynamics is important to understand the different mechanisms which construct the contact network. It is very interesting to note that it depends on the $C_{SE}(\xi, t)$ dynamics and on the triplet configuration $T_{SII}(\xi, \xi')$.

4 Mean Field Approximation

In the first moment approximation, we consider that the spatial structure is completely uniform and that contacts between susceptible and infectious individuals are uniform too. In this case, we can express the second moment as a function of the first moments: $C_{SI}(\xi) = N_S N_I$. Informations about the spatial structure are then lost. In this case, the equations of the first moments are simplified:

$$\frac{dN_S}{dt} = -\beta N_S N_I \quad (12)$$

$$\frac{dN_E}{dt} = \beta N_S N_I - \alpha N_E \quad (13)$$

$$\frac{dN_I}{dt} = \alpha N_E - \gamma N_I \quad (14)$$

It is worthwhile noticing that this system is the same as the most classic mean field model used in epidemiology, usually called SEIR compartmental model. It

has been widely developed and is still of a high interest of study [28]. In this model, the number of individuals is stationary and the space is not represented. The force of infection $\lambda(I) = \beta N_I$ is called density dependent because it only depends on the infectious individual mean density N_I .

In this approximation method, contact structure is implicitly taken into account through pair correlation densities. However, it totally disappears in the mean field approximation as we explain it in the next paragraphs.

For the same reasons as explained section 2.2 we can write the instantaneous infection probability for any susceptible individual as follows (as shown in [26, box 2.1]):

$$I = \log(1 - c) \cdot k_I \quad (15)$$

where k_I is the expected contact rate with infectious individuals and "c" is the probability to be infected if a contact occurs. Equation 12 is the computation of the expected number of infectious individuals during "dt" and can be written as:

$$\frac{dN_S}{dt} = I \times N_S \quad (16)$$

Hence, when we use equation 12, we assume that for each susceptible individual,

$$k_I = k \cdot N_I$$

where $k = 1 \text{ contact} \cdot s^{-1} \cdot \text{surface} \cdot \text{individual}^{-1}$ is a constant contact rate per unit of infectious individual density. In other words, using the mean field model, we assume that the contact structure is a complete graph which all links values are equals (the value is: $\frac{k}{\text{surface}}$). Consequently, there is no contact structure.

Note that we use $\beta = \log(1 - c)$ different from the convention $\beta = k \cdot \log(1 - c)$ because we consider $k = 1 \frac{\text{contact}}{s \cdot \text{individual} \cdot \text{surface}}$.

5 Simulations

5.1 Experimental plan

The purpose of our experimental plan is to compare the results of the simulations of the IBM, MF and MA models. Comparison between IBM and MF results tells us if the spatial pattern plays a role in the dynamics. We have seen in the model presentations that MF considers an homogeneous spatial structure and that the MA simulates a part of the spatial structure. On the opposite, the IBM is our reference and simulates the complete structure of contacts. So if MA results are closer to IBM results than MF results are, we can assume that MA succeed to incorporate pattern dynamics. To do that, we need to define a comparison criterion. We first define f_j as:

$$f_j = \int N_S(N_E) dN_E + \int N_S(N_I) dN_I + \int N_I(N_E) dN_E \quad (17)$$

where $j = IBM, MA$ or MF . It calculates the integrands of the dynamics. Then, the following function $f_{j,k}^{dif}$ (where $k = IBM, MA$ or MF , and $j \neq k$) is used to evaluate the difference between two types of simulations:

$$f_{j,k}^{dif} = |f_j - f_k| \quad (18)$$

This criterion has been chosen because it is independent of the time t . Indeed, we have observed a shift of the dynamics with respect to the time. This shift is due to the fact that spreads don't begin exactly at the same time in the IBM simulations, especially for low values of β . Note that the dynamics of N_R is not considered in order to have a criterion based on three independent variables (N_R can be deduced from the other dynamics and is redundant). The value of $f_{j,k}^{dif}$ is calculated till the system is considered as stationary. The system is considered as stationary when the number of exposed individuals and the number of infectious individuals are equal to 0.

As we said in the introduction our example is Highly Pathogenic H5N1 avian influenza virus in a chicken flock. Concerning initial conditions, we consider that we are in an endemic situation so several individuals are already infectious when the simulation starts. Following paragraph describe the parameter values we choose.

For parameter γ , we choose an interval observed during laboratory experiments for this virus ([29]), corresponding to an infectious period ($\frac{1}{\gamma}$) between 12 hours and 48 hours with a 1 minute time step. For parameter α we choose a lower bound corresponding to the mean latency period ($\frac{1}{\alpha}$) observed for this virus (corresponding to a 6 hours period [29]) and an upper bound corresponding to no latency period, in order to evaluate the impact of latency period on the quality of the MA results. For parameters β , let's notice that they incorporate many characteristics of the system under study and thus there isn't a single value of β that only depends on the virus H5N1. Actually, such transmission parameters are difficult to estimate and are usually calibrated to fit prevalence data observed on the field. For this reason, we choose large intervals for these parameters. To test these intervals, we used the IBM to simulate outbreaks by changing initial conditions to 1 infectious individual and 9999 susceptible individuals. Indeed, we wanted to make sure that chosen intervals for β and λ kept us in realistic situations regarding to H5N1 outbreaks observed on the field in poultry flocks [30, 31]. We verified that these intervals kept us in situations where we observed from 90% mortality in 6 days (worse cases observed on the field [30, 31]) up to situations where the epidemics does not start. Note that in the following, five simulations have been done for the IBM and only the mean result is reported. Note that we have chosen an epidemic situation in which we have several infected individuals at the beginning of the simulation (1000 infectious individuals for 9000 susceptible individuals), so that variability is low and even if we performed only five replicates for each set of parameter values, the differences between the aggregated models (MA and MF) and the IBM are not due to the stochasticity of the IBM.

In the end, we tested four values taken with a regular step in the following

intervals:

- β definition interval is $[0.001, 0.1]$ (tested values are: 0.001, 0.034, 0.067, 0.1);
- α definition interval is $[1/360, 1]$ (tested values are: 0.0028, 0.3352, 0.6676, 1);
- γ definition interval is $[1/720, 1/2880]$ (tested values are: $3.472e^{-4}$, $6.944e^{-4}$, $1.042e^{-3}$, $1.388e^{-3}$);
- λ (move probability) definition interval is $[0.1, 1]$ (tested values are 0.1, 0.4, 0.7, 1).

In order to analyze the moment approximation and the influence of the pattern, we have used a population of 10000 individuals. The grid definition of the IBM is equal to $[150 \times 150]$. A convergence study has been done on the grid size and this value is sufficient in order to have a sufficient discretization of the IBM. The configuration consists of a random distribution of the individuals in an endemic situation. 9000 individuals S and 1000 individuals I have been randomly distributed.

5.2 Influence of the parameters on the spatial structure

Figure 2 presents the difference between the IBM results and the mean field results in order to study the influence of the parameters on the spatial structure. Comments are:

- influence of the latency (α) and the infection (β) : these parameters control the transition of state of an individual from a susceptible to an infectious individual. For high values of α and β , susceptible individual rapidly becomes an infectious individual. It leads to infected cluster of individuals. In this case, we have a large difference with the mean field approximation where the population is supposed to be uniformly distributed. However, the values of β are too low (in our case) in order to have this configuration. On the other hand, for low values of α and β , the correlation values of individuals are lower than the expected values of a uniform distribution. It finally leads to difference with the MF approximation, which considers a uniform distribution. The last example of section 5.3 illustrates this difference regarding to the pair correlations of individuals during the IBM simulation;
- for $\beta = 0.01$, the difference is very important. It is due to the fact that the disease spread doesn't begin in several simulations of the IBM due to low values of β ;
- influence of λ : the higher the value of λ , the higher the mobility of the individuals. This parameter has an important influence on the spatial structure. Indeed, for high values of λ , the mobility is important and the population distribution is very close to a uniform distribution. In this case,

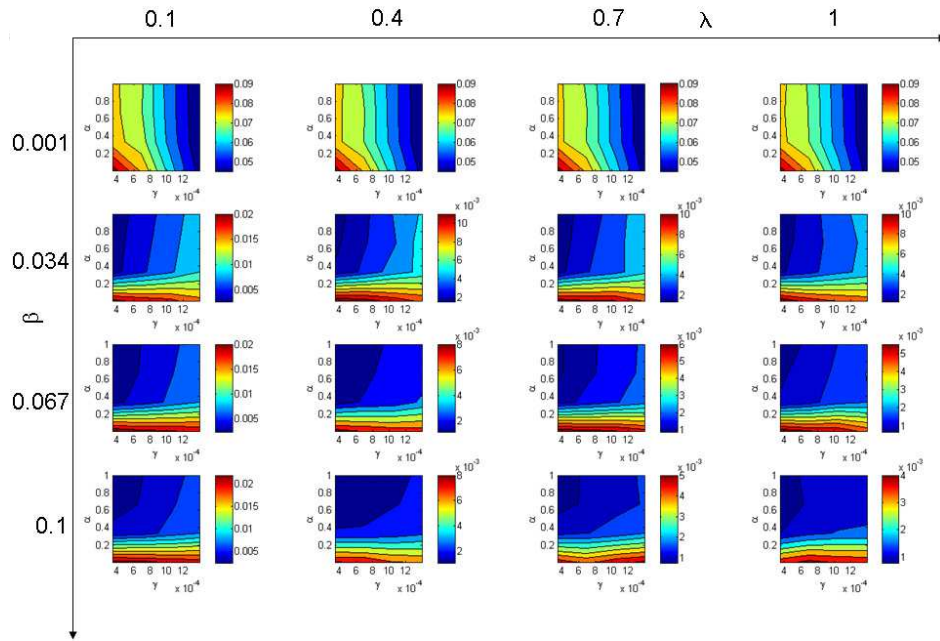


Figure 2: $f_{IBM,MF}^{dif}$ between the IBM and the mean field. (Online version in colour.)

it decreases the difference with the MF approximation; It is noticeable on the first column for $\lambda = 0.1$ where the difference is really important;

- influence of γ : classical values of γ are low. It leads to a weak influence of γ except in the case of low values of β and α . In this case, low values of γ lead to lower correlation values between S and I individuals than the expected values of a uniform distribution and increase the difference with the mean field approximation.

Note that the above comments are only available for the tested values of the parameters (see section 5.1).

5.3 Comparison of the moment approximation with the IBM

Figure 3 presents the difference between the IBM results and the moment approximation results in order to study the validity of the approximation.

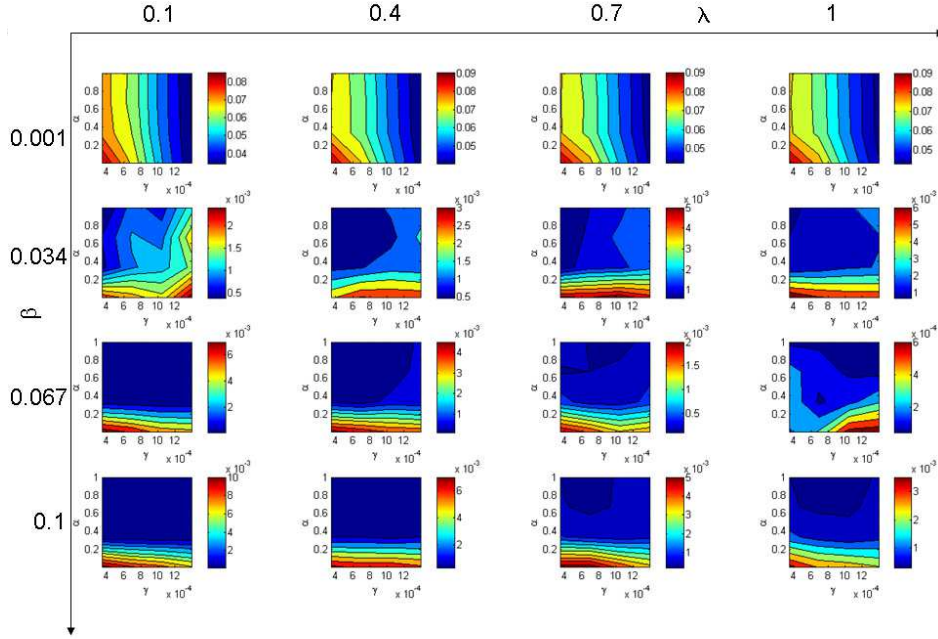


Figure 3: $f_{IBM,MA}^{dif}$ between the IBM and the moment approximation. (Online version in colour.)

Comments are:

- the distribution of $f_{IBM,MA}^{dif}$ is qualitatively very close to the distribution of $f_{IBM,MF}^{dif}$. It can be explained by the fact that the moment approximation simulates the spatial structure dynamics but when the spatial structure becomes somehow too complex, the moment approximation doesn't fully capture the spatial structure. Indeed, we have a second moment approximation and it is not sufficient in the case of complex spatial structures;
- the difference between the IBM and the MA is lower than the difference between the IBM and the MF. The moment approximation gives a better approximation of the dynamics than the mean field approximation because the MA takes into account the spatial structures;
- The influence of the moving rate λ is the same as for $f_{IBM,MF}^{dif}$. The accuracy of the approximations increase with the rate of movement.
- in order to compare the three approaches, dynamics have been plotted on Figure 4 in the case of a spatial structure where the difference is important between the MA and the IBM. Parameters values are $\beta = 0.1$, $\alpha = 1/360$, $\gamma = 3.472e^{-4}$ and $\lambda = 0.1$. We can note that the MA is very close to the IBM results and that the MF is not sufficient to approximate the IBM. Note that this figure shows the case where the difference is the highest between the IBM and the MA. In other case, dynamics of the IBM and of the MA are superimposed. For this case (same parameters and initial conditions), we choose a simulation of the IBM and plotted the (SI) -pair correlation density evolution during the simulation (see Figure 5). Densities are normalized. It means that a density of 1 for a given distance implies that the density of pairs at this distance is the expected density if the population is homogeneously distributed. We can use figure 5 to illustrate figure 4. Figure 5 depicts the evolution of the SI pairs correlation densities of individuals in the IBM. It shows that the average distance between susceptible individuals and infectious individuals is increased at some moment of the simulation. We can observe a clear pattern for the column that corresponds to the 4th hour for instance: the smallest densities are observed for the smallest distances. This means that a spatial structure appeared during the simulation. This spatial structure is responsible of the difference observed between curves of Figure 4. The longer average distance between infectious and susceptible individuals leads to a smaller force of infection. This phenomenon cannot be reproduced by the mean field model and that is the reason why the MF overestimates the number of infectious individuals. Note that when the simulation goes further the spatial structure tends toward a uniform distribution.

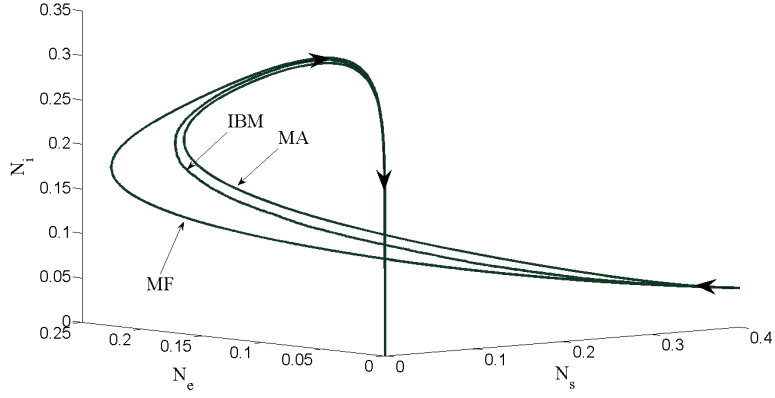


Figure 4: Plane of phases of the IBM, the MA and the MF simulations, for $\beta = 0.1$, $\alpha = 1/360$, $\gamma = 3.472e^{-4}$ and $\lambda = 0.1$.

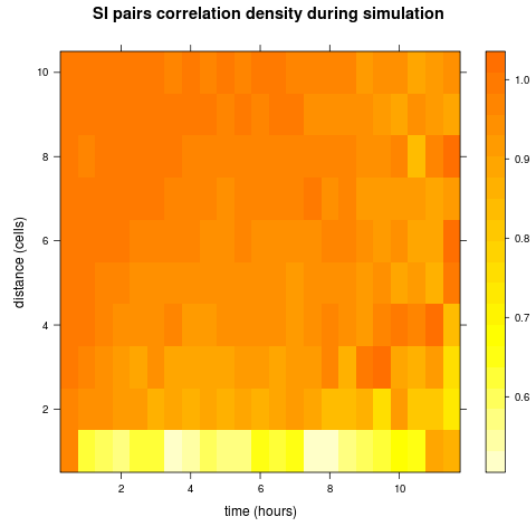


Figure 5: Dynamics of the normalized pair correlation densities in the IBM for simulation of previous figure. Pair correlation density (color code) is computed for distances from 1 cell to 10 cells each simulated half hour (30 time steps) until there is 10 susceptible individuals left (12 hours here).

6 Discussion and Conclusion

In this paper we used Moment Approximation technique to address the issue of simulating epidemic spread with aggregated models when the structure of contacts between individuals is dynamic and complex. To do that, we choose the example of an epidemic system with moving hosts: the spread of the highly pathogenic avian influenza virus H5N1 in a poultry flock.

We first presented an individual based model of this system and we used this IBM as a reference. Then we built two approximated models, the Moment Approximation model, and the Mean Field approximation model.

We compared the simulations of the three models for different values of parameters chosen through a regular exploration of the parameter space. As we develop it in the models presentation, the MF model does not consider any spatial structure. We have seen indeed that the mean field approximation corresponds to a moment closure at the second moment. In the MF the same probability of contact is considered for any pair of individuals. In the opposite, in the IBM, the probability of contact for a pair of individuals depends on the distance between the individuals which depends on the system history. So the comparison between IBM results and MF results gave us a measurement of the influence of spatial structure on the disease spread dynamics for different values of the parameters.

To build the MA model, we used a moment approximation closed at the third moment. This means that the dynamics of the second moment is simulated. In other words, we take into account the mean distances between pairs of different types of individuals. The approximation is performed for the third moment (see appendix B.1). So a part of the spatial structure is taken into account and we observe indeed that the simulation results of the MA are closer to the IBM results than MF results are. We also observe that MA results and IBM results are very close and we then conclude that MA integrates correctly the dynamic spatial structure.

Further work would be necessary to define precisely a criterion that would enable to discriminate when the mean field model is sufficient, when the moment approximation is necessary and sufficient, and when only an IBM is able to describe the dynamics. We think that such criterion should be defined according to the modelling objectives. The most important feature of an epidemic model is not always to give the best prediction of the evolution of the prevalence. It can be for instance to establish a relation between the mortality observed, and the number of infectious individuals in the population. It can also be a qualitative output, that tells for instance if the disease will become endemic in the population or if it will ends after an epidemics. This kind of behaviour can be observed for instance in the case of models with reintroductions of susceptible individuals. The system can fall in an equilibrium state where the number of infectious individuals never reaches zero.

However, this first study shows that in cases where we cannot depict the contact network, MA method provides indeed a compromise between the simulation of every individual move and the simulation of a mean field model which

would ignore any structure in the population.

Lack of information about contacts between animals has already led modellers to build IBM based on animal behaviours provided by Ethologists. Then, they use the outputs of the IBM to calibrate a metamodel that reproduces IBM dynamics. The metamodel is a mathematical function that approximates the model response, has a negligible computation cost, and can predict accurately new responses [25]. A complete and applied example of such study is presented in [16] and [32]. In these papers, the authors first presented an IBM at the herd level [16] and then they used a metamodel to simulate the herd level model in a wider model [32]. The difference between these works and the method we offer in this paper is that in our case we build the approximated model based on the processes modeled and not based on the outputs obtained as it is the case when modelers calibrate metamodels with model outputs. We can see in our discussion that the differences between the models and the approximated models can be explained based on the level of complexity considered by the approximation.

With this study, we want to emphasize the advantages to use moment approximation method versus metamodeling methods based on calibration. We can see in this paper that moment approximation method force the modeller to focus on the processes driving the system. In our case, it gives him an explicit formulation of contact dynamics. We think that in every cases where such approximation is possible, models obtained with moment approximation should be preferred to calibrated metamodels because of their transparency feature. In this paper, our approach to build the moment approximation was to transfer the individual rules to pair-wise rules, and then to triplet-wise rules. In future work, it would be interesting to asses whether if the same approach can be used for existing IBM with moving individuals that are used in epidemiology. We think that such approach would be successful as far as the IBM is clearly specified. However, tests remain to be done to asses which kind of individual rules (such as complex logical rules) would make it impossible for us to use the moment approximation method.

Apart from improving understanding and precision, we think that simulating some features of the spatial patterns are valuable. These features are in our case the distance distribution between pairs of different types of individuals. It has been shown that these features can be used to evaluate models [33] and thus to better calibrate it. Moreover, new methods are developed to infer epidemics dynamics from this kind of data in the case of static individuals [34]. We think that if we can simulate the second moment successfully we will be able to use the same methods in the case of moving individuals.

To conclude, we think that this study is interesting for two main reasons. Firstly, it highlights the fact that moment approximation techniques are not the prerogative of networks-based models: this technique can be used when the network is unavailable and moreover, it enables to take into account a dynamic contact structure. Secondly, it gives a complete example of moment approximation in the field of epidemiology.

References

- [1] Shirley, M.D. and Rushton, S.P., 2005 The impacts of network topology on disease spread. *Ecological Complexity* **2(3)**, 287 – 299, ISSN 1476-945X, doi:DOI:10.1016/j.ecocom.2005.04.005.
- [2] Pautasso, M. and Jeger, M.J., 2008 Epidemic threshold and network structure: The interplay of probability of transmission and of persistence in small-size directed networks. *Ecological Complexity* **5(1)**, 1 – 8, ISSN 1476-945X, doi:DOI:10.1016/j.ecocom.2007.07.001.
- [3] Schimit, P. and Monteiro, L., 2009 On the basic reproduction number and the topological properties of the contact network: An epidemiological study in mainly locally connected cellular automata. *Ecological Modelling* **220(7)**, 1034 – 1042, ISSN 0304-3800, doi:DOI:10.1016/j.ecolmodel.2009.01.014.
- [4] Fuentes, M.A. and Kuperman, M.N., 1999 Cellular automata and epidemiological models with spatial dependence. *Physica A: Statistical and Theoretical Physics* **267(3-4)**, 471 – 486, ISSN 0378-4371, doi:DOI:10.1016/S0378-4371(99)00027-8.
- [5] Kuperman, M. and Wio, H., 1999 Front propagation in epidemiological models with spatial dependence. *Physica A: Statistical Mechanics and its Applications* **272(1-2)**, 206 – 222, ISSN 0378-4371, doi:DOI:10.1016/S0378-4371(99)00284-8.
- [6] Barthélemy, M., Barrat, A., Pastor-Satorras, R. and Vespignani, A., 2005 Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Journal of Theoretical Biology* **235(2)**, 275 – 288, ISSN 0022-5193, doi:DOI:10.1016/j.jtbi.2005.01.011.
- [7] Keeling, M., 2005 The implications of network structure for epidemic dynamics. *Theoretical Population Biology* **67(1)**, 1 – 8, ISSN 0040-5809, doi:DOI:10.1016/j.tpb.2004.08.002.
- [8] Turechek, W.W. and Madden, L.V., 2001 Effect of scale on plant disease incidence and heterogeneity in a spatial hierarchy. *Ecological Modelling* **144(1)**, 77 – 95, ISSN 0304-3800, doi:DOI:10.1016/S0304-3800(01)00350-7.
- [9] Keeling, M.J. and Eames, K.T., 2005 Networks and epidemic models. *Journal of The Royal Society Interface* **2(4)**, 295–307, doi:10.1098/rsif.2005.0051.
- [10] Franc, A., 2004 Metapopulation dynamics as a contact process on a graph. *Ecological Complexity* **1(1)**, 49–63, ISSN 1476-945X, doi:DOI:10.1016/j.ecocom.2003.10.002.
- [11] Eames, K., Read, J. and Edmunds, W., 2009 Epidemic prediction and control in weighted networks. *Epidemics* **1(1)**, 70–76.

- [12] Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J. and Grenfell, B.T., 2001 Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. *Science* **294**(5543), 813–817, doi:10.1126/science.1065973.
- [13] House, T. and Keeling, M.J., 2010 Insights from unifying modern approximations to infections on networks. *Journal of The Royal Society Interface* doi:10.1098/rsif.2010.0179.
- [14] Perkins, S.E., Cagnacci, F., Stradiotto, A., Arnoldi, D. and Hudson, P.J., 2009 Comparison of social networks derived from ecological data: implication for inferring infectious disease dynamics. *Journal of Animal Ecology* **78**, 1015–1022, doi:DOI:10.1111/j.1365-2656.2009.01557.x.
- [15] Thulke, H.H., Grimm, V., Müller, M.S., Staubach, C., Tischendorf, L., Wissel, C. and Jeltsch, F., 1999 From pattern to practice: a scaling-down strategy for spatially explicit modelling illustrated by the spread and control of rabies. *Ecological Modelling* **117**(2-3), 179 – 202, ISSN 0304-3800, doi:DOI:10.1016/S0304-3800(98)00198-7.
- [16] Sabatier, P., Durand, B., Dubois, M., Ducrot, C., Calavas, D. and de Wille, A.V., 2004 Multiscale modelling of scrapie epidemiology: I. herd level: a discrete model of disease transmission in a sheep flock. *Ecological Modelling* **180**(2-3), 233–252, ISSN 0304-3800, doi:DOI:10.1016/j.ecolmodel.2004.05.012.
- [17] Amouroux, E., Desvaux, S. and Drogoul, A., 2008 Towards virtual epidemiology: An agent-based approach to the modeling of H5N1 propagation and persistence in north-vietnam. In S.B. Heidelberg (ed.), *Intelligent Agents and Multi-Agent Systems*, vol. Volume 5357/2008, pp. 26–33.
- [18] Grimm, V., Berger, U., DeAngelis, D., Polhill, G., Giske, J. and Railsback, S., 2010 The odd protocol: a review and first update. *Ecological Modelling* **221**, 2760–2768.
- [19] Duboz, R., Versmisse, D., Travers, M., Ramat, E. and Shin, Y., 2010 Application of an evolutionary algorithm to the inverse parameter estimation of an individual-based model. *Ecological Modelling* **221**, 840–849.
- [20] Dieckmann, U. and Law, R., 2000 *The geometry of ecological interactions: simplifying spatial complexity*. Cambridge University Press, Cambridge, UK.
- [21] Law, R., Murrell, D. and Dieckmann, U., 2003 Population growth in space and time: spatial logistic equations. *Ecology* **84**(1), 252 – 262.
- [22] Filipe, J. and Gibson, G., 2001 Comparing approximations to spatio-temporal models for epidemics with local spread. *Bulletin of Mathematical Biology* **63**, 603–624, ISSN 0092-8240, 10.1006/bulm.2001.0234.

- [23] Bolker, B.M., 1999 Analytic models for the patchy spread of plant disease. *Bulletin of Mathematical Biology* **61**(5), 849 – 874, ISSN 0092-8240, doi: DOI:10.1006/bulm.1999.0115.
- [24] Treuil, J.P., Drogoul, A. and Zucker, J.D., 2008 *Modélisation et simulation á base d’agents - Exemples commentés, outils informatiques et questions théoriques*. Dunod, France.
- [25] Kleijnen, 2008 *Design and Analysis of Simulation Experiments*. Springer
- [26] Keeling, M. and Rohani, P., 2007 *Modelling Infectious Diseases*. Princeton University Press.
- [27] Kirkwood, J., 1935 Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **3**(300).
- [28] Choisy, M., Gueguan, J.F. and Rohani, P., 2007 *Encyclopedia of Infectious Diseases: Modern Methodologies*, chap. Mathematical Modeling of Infectious Diseases Dynamics. John Wiley and Sons Inc.
- [29] Bouma, A., Claassen, I., Natih, K., Klinkenberg, D., Donnelly, C.A., Koch, G. and van Boven, M., 01 2009 Estimation of transmission parameters of H5N1 avian influenza virus in chickens. *PLoS Pathog* **5**(1), e1000281, doi:10.1371/journal.ppat.1000281.
- [30] Tiensin, T., Nielen, M., Vernooij, H., Songserm, T., Kalpravidh, W., Chotiprasatintara, S., Chaisingh, A., Wongkasemjit, S., Chanachai, K. *et al.*, 2007 Transmission of the highly pathogenic avian influenza virus h5n1 within flocks during the 2004 epidemic in thailand. *Journal of Infectious Diseases* **296**(11).
- [31] Tsukamoto, K., Imada, T., Tanimura, N., Okamatsu, M., Mase, M., Mizuhara, T., Swayne, D. and Yamaguchi, S., 2007 Impact of different husbandry conditions on contact and airborne transmission of H5N1 highly pathogenic avian influenza virus to chickens. *Avian Diseases* **51**(1), 129–132, doi:10.1637/0005-2086(2007)051[0129:IODHCO]2.0.CO;2.
- [32] Durand, B., Dubois, M.A., Sabatier, P., Calavas, D., Ducrot, C. and de Wielle, A.V., 2004 Multiscale modelling of scrapie epidemiology: II. geographical level: hierarchical transfer of the herd model to the regional disease spread. *Ecological Modelling* **179**(4), 515–531, ISSN 0304-3800, doi: DOI:10.1016/j.ecolmodel.2004.05.015.
- [33] Henebry, G.M., 1995 Spatial model error analysis using autocorrelation indices. *Ecological Modelling* **82**(1), 75 – 91, ISSN 0304-3800, doi:DOI:10.1016/0304-3800(94)00074-R.
- [34] Filipe, J., Otten, W., Gibson, G. and Gilligan, C., 2004 Inferring the dynamics of a spatial epidemic from time-series data. *Bulletin of Mathematical Biology* **66**, 373–391, ISSN 0092-8240, 10.1016/j.bulm.2003.09.002.

Appendices

A IBM discretization

To simulate the IBM, we consider a regular time step Δt and a discrete 2D-space (a square toroidal 8-neighbourhood lattice). We discretized the spatial kernels ω_1 and ω_2 for the move and infection processes (equations 1 and 3). We used an uniform kernel function with a size equal to 1 discrete space step for the infection kernel ω_2 (see equation 19). It leads to consider a contact per time step between two individuals if these individuals are located in a same case during Δt :

$$\begin{aligned}\omega_2([0, 0]) &= 1 \\ \omega_2(x \neq [0, 0]) &= 0\end{aligned}\tag{19}$$

We used an uniform kernel function with a size equal to 3 discrete space steps for the move kernel ω_1 (see equation 20). The individual can only move in a neighbour square:

$$\begin{aligned}\omega_1([x, y]) &= \frac{1}{8} \quad \text{if } x = \pm 1 \text{ or } y = \pm 1 \\ \omega_1([x, y]) &= 0 \quad \text{otherwise}\end{aligned}\tag{20}$$

At the initialisation, a state $\{x_i^0, s_i^0\}$ (with $x_i \in \mathbb{N}^2$ and $s_i \in \{S, E, I, R\}$) is given for each individual $i \in [1, N]$. Then, for each time step, for each individual, we compute the next state $\{x_i^{t+\Delta t}, s_i^{t+\Delta t}\}$ according to the previous global state $(\{x_i^t, s_i^t\})_{1 \leq i \leq N}$.

- **Computing new infectious state $s_i^{t+\Delta t}$**

According if individual infectious state is respectively S , E or I , we perform a random selection using respectively probability P_E , P_I or P_R computed with equations 3, 4 or 5 to compute if infectious state will be changed to respectively E , I or R . If the test succeeds, the state is changed, if not, it stays still.

- **Computing new position $x_i^{t+\Delta t}$**

We first compute if individual moves or stays on the same cell. We perform a random selection of probability $P = P_M(x_i^{t+\Delta t} \neq x_i^t)$. If the random selection fails, position is not changed, if it succeeds, the new position is computed as $x_i^{t+\Delta t} = x_i^t + (x, y)$ with (x, y) randomly chosen in $\{(x, y) \mid x = \pm 1 \text{ and } y = \pm 1\}$

B Development and discretization of the moment approximation

We present here the development of pair correlation dynamics. The pair correlation $C_{SS}(\xi, t)$ can be calculated for the susceptible-susceptible pairs as follows:

$$\begin{aligned} \frac{dC_{SS}(\xi, t)}{dt} = & \\ \text{Movement} \left\{ \begin{aligned} & 2\lambda \int \omega_1(\xi') C_{SS}(\xi + \xi') d\xi' \\ & - 2\lambda |\omega_1(\xi)| C_{SS}(\xi) \end{aligned} \right. & (21) \\ \text{Infection} \left\{ \begin{aligned} & - 2\beta \int \omega_2(\xi') T_{SSI}(\xi, \xi') d\xi' \end{aligned} \right. \end{aligned}$$

The susceptible-susceptible pair correlation dynamics depends on the two processes in the above equation:

- the movement terms (the two first) are computed as the same way as for the C_{SI} pairs (see section 3.2);
- the infection term computes the expected number of susceptible individuals in situation ξ that get infected. It depends on the triplet configuration T_{SSI} and of the interaction kernel ω_2 . If an infectious individual is located at distance ξ' of an individual of a pair at distance ξ , it can destroy this pair.

There is a factor 2 before each term. It is due to the fact that the process can be applied on each individual of the pair.

The pair correlation $C_{SE}(\xi, t)$ can be calculated for the susceptible-exposed pairs:

$$\begin{aligned} \frac{dC_{SE}(\xi, t)}{dt} = & \\ \text{Movement} \left\{ \begin{aligned} & + 2\lambda \int \omega_1(\xi') C_{SE}(\xi + \xi') d\xi' \\ & - 2\lambda |\omega_1(\xi)| C_{SE}(\xi) \end{aligned} \right. & (22) \\ \text{Infection} \left\{ \begin{aligned} & + \beta \int \omega_2(\xi') T_{SEI}(\xi, \xi') d\xi' \\ & - \beta \int \omega_2(\xi') T_{SEI}(\xi, \xi') d\xi' \end{aligned} \right. \\ \text{Latency} \left\{ \begin{aligned} & - \alpha C_{SE}(\xi) \end{aligned} \right. \end{aligned}$$

The susceptible-exposed pair correlation dynamics depends on the three processes in the above equation:

- the movement terms (the two first) are computed as the same way as the C_{SS} pairs;
- the infection terms create and destroy some (SE) -pairs. The creation of a (SE) -pair depends on the T_{SSI} configuration and of the interaction kernel ω_2 . When the susceptible individual of a (SS) -pair is infected by an infectious individual located at distance ξ' , it creates a new (SE) -pair. The destruction of a (SE) -pair depends on the T_{SEI} configuration and of the interaction kernel ω_2 . When the susceptible individual of a (SE) -pair is infected by an infectious individual located at distance ξ' , it destroys the (SE) -pair;
- the latency term corresponds to the number of exposed individuals that are in a configuration distant of ξ with a susceptible individual and hence create a new (SI) -pair distant of ξ if they become infectious.

The dynamics of the second moments depend on the third moment $T_{ijk}(\xi, \xi')$. This moment has to be closed in order to perform the calculation. The moment closure is studied in section B.1.

B.1 Moment closure

To achieve a closed dynamical system, the highest spatial moment must be replaced by a function of lower-order moments. This expression constitutes the moment closure. In our case, the third moment must be replaced by a function of the first and second moment. Moment closures constitute a key-issue for moment approximation because the quality of the moment approximation is directly linked to the used moment closure. Several closures have been tested and studied in the literature [20]. The idea is to use different closures which have been tested by [20]. More details are available in these references, especially for the all properties linked to the different closures. We have used a power-2 closures which have been already used by [20]. Power-2 closures are obtained by multiplying two of the three pair densities and by dividing by the density of the opposite corner. In our case, it is important to take into account the correlation which leads to the infection of susceptible individuals. It leads to the following approximation of the third moment:

$$T_{ijk}(\xi, \xi') = \frac{C_{ij}(\xi)C_{ik}(\xi')}{N_i} \quad (23)$$

If we consider this closure and the third moment used in the approximation, the correlation involved in the infection process is taken into account.

B.2 Discretization

In order to implement the moment approximation, it is necessary to discretized the correlation functions. For this purpose, the same discretization as in the individual-based model is used. The movement kernel and the interaction kernel

are the kernels described in appendix A. An infection of susceptible occurs only if an infectious individual is located in the same case. The individuals move to an adjacent case with a uniform probability. The current problem is considered as bidimensionnal. In other terms, we consider correlation matrix. It leads to a simplification of the dynamics as follows:

$$\frac{dN_S}{dt} = -\beta C_{SI}(0, 0) \quad (24)$$

$$\frac{dN_I}{dt} = \beta C_{SI}(0, 0) - \alpha N_e \quad (25)$$

$$\frac{dN_I}{dt} = \alpha N_e - \gamma N_i \quad (26)$$

for the first moment dynamics. For the second moment dynamics, we have:

$$\begin{aligned} \frac{dC_{SS}(x_i, x_j, t)}{dt} = & \\ & \text{Movement} \left\{ \begin{aligned} & 2\lambda \sum_{i=0}^2 \sum_{j=0}^2 C_{SS}(x_{i-1}, x_{j-1}) \\ & - 2\lambda C_{SS}(x_i, x_j) \end{aligned} \right. \quad (27) \\ & \text{Infection} \{ -2\beta T_{SSI}([x_i, x_j], [0, 0]) \} \end{aligned}$$

$$\begin{aligned} \frac{dC_{SI}(x_i, x_j, t)}{dt} = & \\ & \text{Movement} \left\{ \begin{aligned} & + 2\lambda \sum_{i=0}^2 \sum_{j=0}^2 C_{SI}(x_{i-1}, x_{j-1}) \\ & - 2\lambda C_{SI}(x_i, x_j) \end{aligned} \right. \quad (28) \\ & \text{Infection} \left\{ \begin{aligned} & - \beta T_{SII}([x_i, x_j], [0, 0]) \\ & - \beta C_{SI}(x_i, x_j) \delta(x_i) \delta(x_j) \end{aligned} \right. \\ & \text{Recovery} \{ -\gamma C_{SI}(x_i, x_j) \} \\ & \text{Latency} \{ -\alpha C_{SE}(x_i, x_j) \} \end{aligned}$$

$$\begin{aligned} \frac{dC_{SE}(x_i, x_j, t)}{dt} = & \\ & \text{Movement} \left\{ \begin{aligned} & + 2\lambda \sum_{i=0}^2 \sum_{j=0}^2 C_{SE}(x_{i-1}, x_{j-1}) \\ & - 2\lambda C_{SE}(x_i, x_j) \end{aligned} \right. \quad (29) \\ & \text{Infection} \left\{ \begin{aligned} & + \beta T_{SEI}([x_i, x_j], [0, 0]) \\ & - \beta T_{SEI}([x_i, x_j], [0, 0]) \end{aligned} \right. \\ & \text{Latency} \{ -\alpha C_{SE}(x_i, x_j) \} \end{aligned}$$

Annexe F

Le Modèle de Processus Expérimental : Application à la modélisation multi-échelles

The Experimentation Process Model

Application to multiscale modelling

Bruno Bonté¹, Gauthier Quesnel², Jean-Pierre Muller³, Raphaël Duboz¹

¹CIRAD
UR22 Animal et
Gestion Intégrée des Risques
Campus International de Baillarguet
34398 Montpellier cedex 5, France
bruno.bonte@cirad.fr
phone: + 33 (0)4.67.59.39.86
fax: + 33 (0)4.67.59.37.99

²INRA
UR875 Biométrie et
Intelligence Artificielle,
F-31326 Castanet-Tolosan, France

³CIRAD
UR47 Gestion des Ressources
renouvelables Et ENvironnement
Campus International de Baillarguet
34398 Montpellier cedex 5, France

Keywords: DEVS, Virtual Laboratory Environment, Experimental Frame, Recursive Simulation, Epidemiology

Abstract

Modelling and Simulation become the main experimental tools in natural and social sciences such as Ecology and Epidemiology. Experimental designs and optimization techniques require more and more simulations. In this paper, we propose a formal specification of a dynamical experimentation design. We specify it as a generic dynamic system model called Experimentation Process Model (EPM). The EPM is specified in DEVS, thus it can easily be coupled with other DEVS models. We illustrate the concrete applicability of the EPM with an example of multi-scale modelling in the field of Epidemiology.

1. INTRODUCTION

In scientific fields such as Ecology, Economy or Sociology, Modelling and Simulation activities become more and more important due to the social demand for reliable decisions. These disciplines have in common the impossibility to perform rigorous large scale experiments on the real systems. Therefore, the experimentation process cannot be fully addressed, in particular the need to control experimental conditions to ensure reproducibility. Modelling and Simulation therefore appear as one alternative to cope with such difficulties. We use the "virtual laboratory" metaphor where simulations are considered as virtual experiments [Quesnel et al., 2009] in opposition to real experiments.

Repeatability, distribution and automation of scientific experiments are still improving for real experiments. We can mention recent works for unified experiment specification [Maccagnan et al., 2010], as well as for experiment distribution [Goble et al., 2010]. Virtual experiments are performed on dynamic systems models. Formalisms already exist in order to entirely specify virtual experiments and models. Models are built to study a given question in a given context.

The theory of Modelling and Simulation [Zeigler et al., 2000] gives us a way to specify the model and its context using respectively the Discrete Event System (DEVS) specification and the Experimental Frame concept (EF). The DEVS specification enables to specify models and hence ensures models repeatability and distribution. The EF enables to specify the context in which the system under study is considered, and thus in which the model is to be used. It is built together with the research question we want to address with the model. In the same way as the model, it can be specified at different degrees of details called levels of specification [Traore and Muzy, 2006]. This allows to build web repositories with models and their associated EF [Chreyh and Wainer, 2009]. Worthwhile noting that the Experimental Frame concept has now been introduced in ecological modelling community [Aumann, 2007]. We can notice indeed that we have nowadays at our disposal models and their contexts entirely specified.

Additionally, some particular models need to embed a model experimentation process to compute their own behaviours. It is the case in models containing anticipative agents for instance, where agents have to simulate their own future [Gilmer and Sullivan, 2005]. Actually, a model of anticipative agent includes a function which can simulate a model of its representation of the world. The present work is motivated by the fact that EF and models can both be specified in DEVS. Thus we can use these concepts to specify the virtual experiment of a model (we call this model M_E). We can then create another model (that we call M_B), which will manipulate these virtual experiments. We call this technique "embedded simulation". Let's explain it more precisely before coming back to our main purpose. The model M_B is simulated, which consequently implies two levels of simulations. We will call the simulation of M_B "the base simulation", and the simulation of M_E "the embedded simulation". The model M_B thus launches simulations of model M_E . The figure 1 illustrates this technique. The main steps are the following:

- When needed, the base simulation stops,
- experiments specifications are generated by the base simulation,
- embedded simulations run,
- embedded simulations outputs are analysed,
- results return to the base simulation,
- the base simulation can carry on.

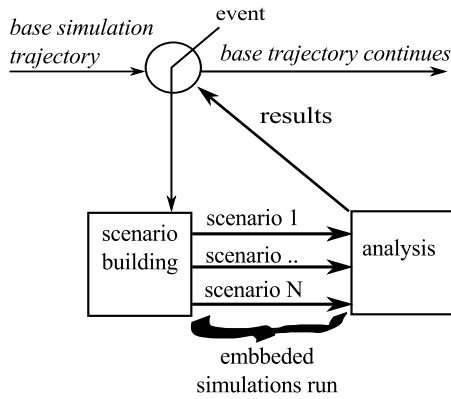


Figure 1. Embedded simulation principle (adapted from Gilmer and Sullivan [2005]). Embedded simulations are launched while a base simulation is running. Embedded-simulations results can then be used to compute values useful for the base simulation. (See section 1.).

Our objective is to develop a method that enables to design and implement the embedded simulation technique understood as virtual experiments within a virtual experiment. We want to propose a framework to specify how virtual experiments are chosen and run in an iterative and dynamic process. Our motivation is to provide a generic methodology to specify, and thus implement, the coupling between the experimentation process and any other dynamic system. Thus, we propose a generic framework to include virtual experimentation process within a simulation. To be operational, this framework is designed as a dynamic system model. We call this model: the Experimentation Process Model (EPM). On the opposit, we will refer to the models simulated in the embedded simulation as the "Experimented Models" (EM). The EPM must be specified and implemented in such a way that it can be itself included as a modular component in any simulation. The EPM is itself a model, so we assume that the most relevant formalism for the EPM specification is the Discrete EVents system Specification (DEVS) [Zeigler et al., 2000]. Indeed, DEVS offers a hierarchical and modular approach that fits well with re-usability. Moreover, as we have specified

this experimentation process model in DEVS, we can use all the integrating properties of DEVS to perform the coupling between the EPM and other models.

Having such an EPM, we can also use it on its own when we want only to implement an experimental process, such as optimisation techniques, sensitivity analysis, system viability studies, etc. However, we do not need the EPM to do that. Many modelling and simulation environments already have features that enable taking parameters and initial states from a global workspace, trigger off simulation runs from this workspace, and then bringing back the simulations results to the workspace. We can mention as examples PowerDEVS with Scilab [Bergero and Kofman, 2011] or VLE (the environment we use) with R statistical software [Quesnel et al., 2009]. These features are very useful and used but they don't provide a proper formalism to couple the simulation runs with another dynamic system model. This is the reason why we need the EPM. The interest of the EPM lies in its relation with the wider model it is included in. The coupled model resulting of the coupling of the EPM and an other dynamic system specified in a DEVS compliant formalism can be itself specified as a DEVS model, thanks to DEVS property of closure under coupling. To fully illustrate the feasibility and interest of the EPM, we had to choose an example where the EPM was coupled with a dynamic system model. We choose an original example in multi-scale modeling in epidemiology.

The paper is composed of two separate parts followed by a general discussion. The first part describes the methodology proposed to design the EPM and its DEVS specification as generic components. Then the second part gives an example of the use of the EPM for multi-scale modeling.

2. EXPERIMENTATION PROCESS MODELLING

2.1. Experimentation process description

We distinguish two levels in our description of an experimentation process. The lowest level is the *experiment* and the highest level is the *experimental design*. The experimental design is actually the way to generate a set of experiments (called experimental plan) and the way to analyse their results.

We can define an experiment as an experimental frame coupled with a model. We will talk about the reasons why we included the model specification within the experiment specification in section 4.. The concept of experimental frame has been described by Zeigler [Zeigler et al., 2000] and then used by different people in different domains [Zeigler et al., 2000, Traore and Muzy, 2006, Aumann, 2007]. Taking into account that the source system represented by a model evolves in some given context (experimental conditions), the experimental frame is a representation of this context.

The EF is of great interest to study the relations between

pairs of system descriptions. It is used, for instance, to test model validity by specifying the circumstances and points of view under which the model must show the same behaviour as the real system. Thus, we can use it for the validation of the models or the verification of the simulators (see [Zeigler et al., 2000] chapter 14), or the selection of models (see System Entity Structure/Model Base framework chapter 18 of [Zeigler et al., 2000] or [Traore and Muzy, 2006]). In our case, we don't use the Experimental Frame to compare two systems but to specify the conditions under which the model is studied. We want indeed to dissociate the model, and the way it is experimented. In this paper we will hence restrict our use of the Experimental Frame to the following description given by Zeigler: "An experimental frame basically consists of three models: a model for generating inputs to the model under study, a model for collecting output data from the model, and a model for controlling the simulation run. We call these three models a generator, an acceptor, and a transducer respectively. The three models are analogous, in their functions, to a signal generator, an oscilloscope, and a spectrum analyser, respectively, in an electronic circuit experiment." [Zeigler et al., 2000]. However, an EF can have a different aspect and it has an important position in the theory and activity of Modeling and Simulation. These issues will further be discussed in the discussion part of the paper.

An experimental frame can be defined on different levels of specification [Traore and Muzy, 2006]. The lowest level only gives the domains of definition of the inputs and outputs of the model. The highest level defines the experimental frame through its implementation, understood as a system of measurement or as a system of observation. The highest level usually contains the three devices described above.

In our work, we only consider the EF on the highest level of specification and we consider the EF specifying only a single experiment. As we will see later in the discussion, a more abstract use of experimental frames can be done. However, this first investigation paper only requires an operational way to specify the context in which the model is experimented. Thus, in our proposal, the experimental frame has the following features:

- The list of input events to send to the model during the simulation. (Note that these events are dated so this list can be viewed as time series).
- The observation policy that tells which and when the variables have to be observed.
- Pre-treatments to apply to these observations in order to get values of interest.
- Possibly some compliance conditions to be checked.
- A simulation stopping condition (usually a simulation duration but it can be different).

As we have said previously, we define an experiment as the combination of an experimental frame and a model. Hence the model itself must be specified. Whatever model one uses, the following elements can be distinguished: *i. the parameters*, *ii. the initial state* and the rest of the model specification that we will refer to as *iii. the model dynamics*.

We now can define an experimental design being composed of:

- The description of an experimental plan.
- The way to analyse experiment results.
- A procedure to generate a new appropriate experimental plan.

The experimental design can indeed be iterative if the result of an experiment is used to define the next experiment. It is the case for simulation-based optimization scheme where model parameters of an experiment are evaluated according to the results of previous experiments. In this case, it is mandatory to be able to build a new experimental plan and to introduce a feedback between experimental plan results and experimental plan generation process. Figure 2 summarizes such iterative experimental design.

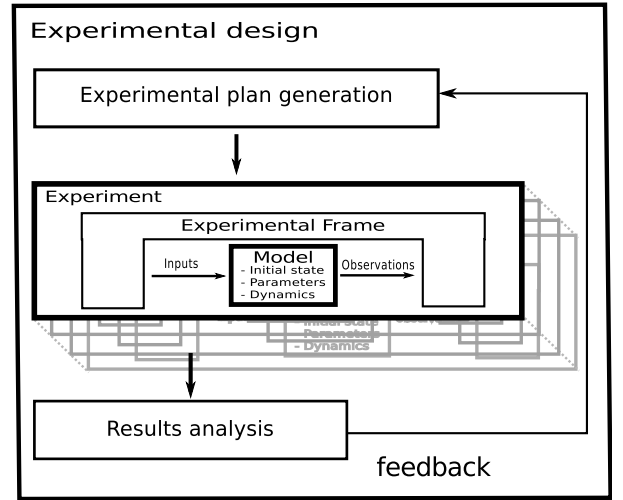


Figure 2. Iterative experimental design. Describes how to generate simulations to run and how to analyse their results.

2.2. Experimentation process modelling using DEVS

In this section, we propose a specification of the EPM. As we said in the introduction, we use the DEVS framework [Zeigler et al., 2000]). The DEVS specification models a dynamic system as a structure $\langle S, X, Y, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda, ta \rangle$ where, S is the set of states, X and Y are respectively the sets

of input and output events, δ_{ext} is the external transition function which describes the state transition when an input event is received, δ_{int} is the internal transition function which describes the state transition, ta is the time advance function to compute the lifespan of the current state, δ_{con} is the confluent function which describes the state transition when the state duration is over and an input event arrives at the same date, and λ is the output function telling what output to send when leaving a state. Such a structure is called an atomic DEVS model and several atomic DEVS models can be coupled to constitute a coupled DEVS model. It is worthwhile noting that both the experimented model and the experimentation process model are specified with DEVS.

2.2.1. Experimentation process coupled model

As we said in the introduction, we want to build a DEVS model of the experimentation process which we will be able to couple with other models. This model is a DEVS specification of the iterative experimental design we presented previously (section 2.1.). Doing that we can specify a dynamics to the iterative experimental design. For sake of modularity, we built it as a coupled model. The coupled model structure is composed of three DEVS models linked as presented figure 3. Next paragraphs present these three models.

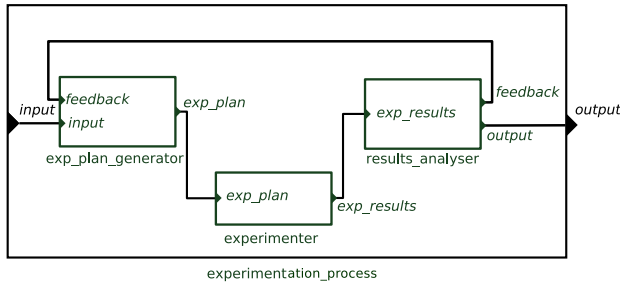


Figure 3. DEVS experimentation process coupled model structure. Picture taken from the VLE software.

The *exp_plan_generator* model has two input ports. The first one receives events from models coupled to the *experimentation_process* model. The second one receives feedback from the *results_analyser_model*. Using information from these ports, it must have functionalities to build an experimental plan (see section 2.1.) and to serialize it in order to send it on its *exp_plan* output port.

The *experimenter* model receives the experimental plan on its input port. The experimental plan is a data structure that contains all necessary information to perform simulations. All the simulations are computed and the results of the experiments are saved in another data structure called experiment results. This data structure is serialized and then sent to its output port.

The *results_analyser* model receives the experiment results data structure and can compute both outcome measures to send out of the *experimentation_process* model, or feedback measures to send back to the *exp_plan_generator* model to initiate a new experimental plan.

Both, experimental plan generator model and result analyser model, are model dependent. Thus we only give their roles and main characteristics in this part. The next part (see section 3.) presents an instantiation of these models as an example. On the opposite, the experimenter model is generic and can be entirely specified.

2.2.2. Experimenter model DEVS specification

We first need to define the following data structures and function.

- *EP* (Experimental Plan): A data structure that contains a list of experiments. These experiments contain each one the following variables:
 - Simulation duration,
 - DEVS model structure,
 - Model dynamics for each DEVS atomic model,
 - Initial state for each DEVS atomic model,
 - Parameters values for each DEVS atomic model,
 - Observed port and frequency of observation for each DEVS atomic model,
 - Number of replicates to runs.
- *ER* (Experiment Results): a structure that contains the outputs of a set of simulations, organized as follows.
 - DEVS sub-model and port from where the observation comes from,
 - Date of the observation,
 - Experiment index,
 - Replicate index.
- *Eval* : $EP \mapsto ER$: The evaluation function uses a DEVS simulation engine to perform all experiments described in an *EP* data structure and fill in an *ER* data structure. More details are given in the implementation description (see section 3.4.4.).

We can now specify the experimenter model as a *DEVS* model structure $\langle X, Y, S, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda, ta \rangle$. Its whole dynamics is illustrated in figure 4, by a DEVS state machine diagram and specified as follows:

- S , the set of states is the following:

$$S = \{(ER, phase, T_0) | phase \in \{INIT, WAIT, EXP\}\}$$

Where ER is an experiment results data structure as described above and T_0 is a time.

- X , the set of input events is the set of pairs of a time value T_i and an experimental plan data structure (EP).
- Y , the set of outputs events is the set of possible experiment results data structure (ER).
- δ_{ext} , the external transition is the following:

$$\delta_{ext} : Q \times X \rightarrow S, \text{ with } Q = \{(s, e) | s \in S, 0 \leq e \leq ta(s)\}$$

$$(EP, T_i), (ER, WAIT, T_0) \mapsto (eval(EP), EXP, T_i)$$

Where $eval()$ is the evaluation function that computes experiment results data structure from experimental plan data structure.

- ta , the time advance function is the following:

$$ta : S \rightarrow \mathbb{R}_0^+ \cup \infty$$

$$(ER, INIT, T_0) \mapsto 0$$

$$(ER, WAIT, T_0) \mapsto \infty$$

$$(ER, EXP, T_0) \mapsto T_0$$

- δ_{int} , the internal transition is the following:

$$\delta_{int} : S \rightarrow S$$

$$\forall (ER, phase, T_0) \in S$$

$$(ER, phase, T_0) \mapsto (ER, WAIT, T_0)$$

- λ , the output function is:

$$\lambda : S \rightarrow Y$$

$$(ER, EXP, T_0) \mapsto ER$$

$$\forall phase \neq EXP, (ER, phase, T_0) \mapsto \emptyset$$

- δ_{con} , the confluent transition is:

$$\delta_{con}(s, x) = \delta_{int}(\delta_{ext}(s, ta(s), x))$$

As we said previously, the experimenter model is generic and can be included as it is in any experimentation process model. Once implemented, one just needs to plug it and it will manage all embedded simulations. It is worthwhile noting some properties of this experimenter model. First, the experimenter model has no endogenous behaviour and will perform experiments only if it is perturbed by an external event.

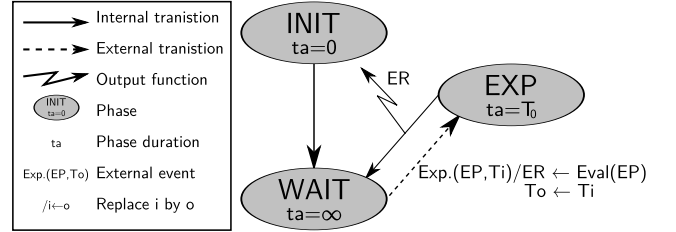


Figure 4. Experimenter model DEVS state machine: phases dynamics.

Second, during this perturbation the experimenter model receives two pieces of information: the experimental plan data structure EP , and the duration $T_0 \geq 0$ of the experimentation phase EXP . Hence, the fact that performing an experimental plan may not be instantaneous can be represented. Let's recall the case of anticipative agents. Agents may need a "thinking period" to perform simulations "in their heads". The duration of this "thinking period" can be specified by T_0 .

3. USE OF THE EPM FOR MULTISCALE MODELLING AND SIMULATION

In this second part, we use the EPM to build a multiscale model. First, we want to illustrate the fact that the experimental process model can be used in a wider model. Secondly, we want to propose an operational framework for the simulation of a particular class of multiscale models¹. As an example, we build a multiscale model in epidemiology.

3.1. Multiscale modelling and simulation

Our work on multiscale modelling is inspired from [Duboz et al., 2003] where multiscale modelling consists in the modelling of a single target system with several models, each one considering a particular time and/or spatial resolution of the target system. The need for multiscale modelling comes from issues related to the understanding of complex systems dynamics, where the microscopic levels of a particular system interacts with the macroscopic levels and *vice versa*. We call scale transfer the way we link different scales (states and processes).

3.2. Related works in ecology and epidemiology

In order to consider the interactions between two scales, it is a commonly used approach in ecology and epidemiology to parametrize macroscopic models using microscopic models [Gianni and Sara, 2007, Jacek and Wojciech, 1999,

¹Work on multiscale modelling is inspired from [Duboz et al., 2003] and from discussions taking place within the Virtual Laboratory Environment (VLE) development team. It has been presented in the SCSC'09 conference Bonté et al. [2009]. Simple models using recursive simulations are available on VLE web site: <http://www.vle-project.org>.

Duboz et al., 2003]. In [Sabatier et al., 2004, Durand et al., 2004] for example, the authors first define a model at a lower scale [Sabatier et al., 2004], and then build a parametric meta-model of this model which is used in a second model defined at an upper scale [Durand et al., 2004]. Such parametrization can have a less straightforward purpose like the determination of the level of details required to reproduce specific functions of an ecosystem [Picard and Franc, 2001]. Hence, several methods coming from other research fields are already used in ecology, such as linear regression, non-linear methods based on least square principle, or mean field approximation, which are the ones used in the quoted papers. What the EPM brings to this kind of work is that it gives a way to dynamically link the microscopic model and the macroscopic model. We develop it hereafter.

3.3. Principle

We want to show that the EPM and associated framework enables to perform simulations using at the same time two models defined at two different spatio-temporal scales. The first scale characterizes slow processes occurring at the macroscopic level and the second scale characterizes fast processes occurring at the microscopic level.

We interpret the interactions between the two levels in the following way:

- The processes described at the macroscopic level determine the context for the processes described at the microscopic level.
- Some properties viewed as simple functions at the macroscopic level are the consequences of complex processes occurring at the microscopic level.

Considering these interactions, we use the EPM to perform the scale transfer. The EPM is included in the macroscopic model simulation and coupled with the macroscopic model. It performs experiments on the microscopic model at run time in order to evaluate aggregated variables of the macroscopic model dependent of microscopic model behaviour (see figure 5).

3.4. Multiscale model in epidemiology

In our example, we use two epidemiological models, an individual based model for the microscopic level, and an aggregated model of disease diffusion for the macroscopic model. We don't insist on scale transfer issues rather than on the method we use to achieve the multi-scale modelling and simulation.

3.4.1. Microscopic Model: IBM

The Individual Based Model (IBM) represents the spread of a disease in a population of individuals moving in a space.

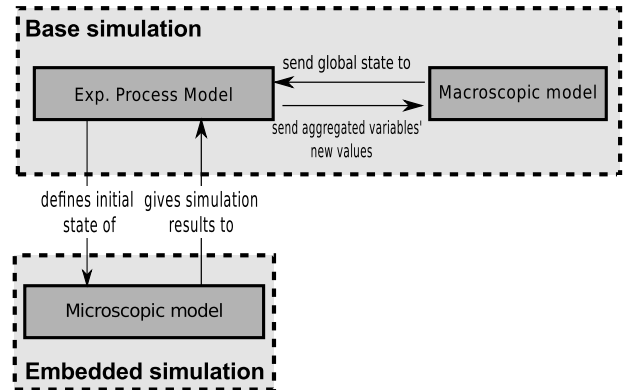


Figure 5. Use of the EPM for scale transfer.

The model is specified within the DEVS formalism and associated extension as proposed by Duboz et al. [2006]. It is composed of:

- An explicit space (a lattice of cells) formalized as a Cell-DEVS [Wainer and Giambiasi, 2001];
- A set of individuals each one formalized as an atomic DEVS model;
- A model of individual movements on the space formalized as a DS-DEVS [Barros, 1996].

We invite the reader to refer to the literature we cite for the definition of space and movement. There is no need here to give the formal definition of individuals' DEVS model but we can give its main features. The state of an individual model contains an "infection state" (S, I or R) standing respectively for Susceptible, Infectious and Recovered individuals; an infectious period (corresponding to the time spent in the infectious state "I", before switching to "R"); and an individual time step. Its dynamics consists in moving at each time step and to change its infectious state. Firstly, the infectious state is switched from "S" to "I" (with a fixed infection probability) if it is on the same cell as an individual of state "I". Then, it is switched from "I" to "R" if it has been infected since the duration of its infectious period. The initial conditions are defined as the number of susceptible, infectious and recovered individuals. At initialisation, the individuals are randomly distributed over the space.

3.4.2. Macroscopic model: SIR

The SIR model is formally described by the following system of differential equations defining the dynamics of three state's variables (equations 1, 2 and 3).

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS & (1) \\ \frac{dI}{dt} &= \beta IS - \gamma I & (2) \\ \frac{dR}{dt} &= \gamma I & (3)\end{aligned}$$

Where:

- S is the density of susceptible individuals;
- I is the density of infectious individuals;
- R is the density of recovered (or removed) individuals;
- β is a constant parameter reflecting the combined effect of all the processes affecting the transmission rate;
- γ is the recovering rate.

The SIR model is one of the simplest model used in epidemiology. It has been widely developed and is still of a high interest of study and used in more complex forms [Choisy et al., 2007]. In this model, the number of individuals is stationary. The individuals are supposed to be uniformly distributed and to move and interact within a continuous toroidal space. Let c and t be respectively arbitrary surface unit and time unit. Then the unit of S , I and R is *individual.c*⁻¹, the unit of β is *individual*⁻¹.*c.t*⁻¹, and the unit of γ is *t*⁻¹.

The SIR model was specified as a single DEVS atomic model using a Quantized State System integrator so that it can be perturbed with very low error propagation [Kofman and Junco, 2001].

3.4.3. Scale transfer EPM

In epidemiology, the infection rate, proportion of susceptible individuals which become infected by time unit, is called the force of infection and often represented by the letter λ . In the simplest way, it can be approximated by the following expression [Choisy et al., 2007]:

$$\lambda = \beta \times I$$

where β is a fixed parameter and I is the number of infectious individuals in the population. The force of infection is represented this way in the SIR model (see equations 1, 2 and 3 section 3.4.2.).

To perform the scale transfer, we consider the parameter β of the SIR model as a variable computed at run time with the IBM. It allows us to evaluate the mean parameter β of the force of infection of a population of individuals having the same behaviours as described in the IBM.

For a single simulation k of the IBM, we estimate the value of β as follows:

Table 1. Scale transfer summary

properties	IBM scale	SIR model scale
simulated duration	t_1	t_2 (with $t_1 < t_2$)
space unit	c	c
space size	a	<i>undefined</i>
population density	d	d

$$\beta_k = \frac{1}{\bar{i}_k \times \bar{s}_k} \times \frac{\Delta s_k}{t} \quad (4)$$

where:

- β_k is the estimated value of β for the simulation k ;
- t is the duration of the simulation;
- \bar{s}_k and \bar{i}_k stand respectively for the mean number of individuals in state s and i over the duration of the embedded-simulation $([0, t])$;
- Δs_k stands for the difference of the number of susceptible individuals between the beginning and the end of the embedded-simulation k .

Then, considering that the IBM is stochastic, the expected value of β is computed as the mean value of β_k over several simulations (n simulations) as in equation 5.

$$\beta = \frac{1}{n} \sum_{k=1}^n \beta_k \quad (5)$$

The IBM has the same density $d = \frac{nb \text{ individuals}}{\text{surface}}$ as the SIR model but in the SIR model the space is not explicitly represented while the IBM is simulated on an area a expressed in the unit cell (" c "). Table 1 summarizes the scale transfer to perform.

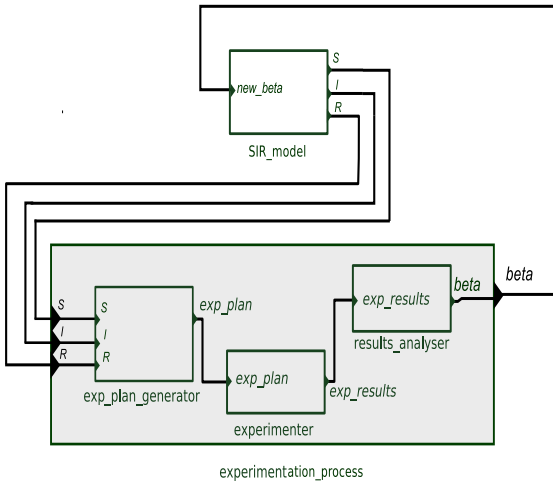
Figure 6 presents how the SIR model (section 3.4.2.) is coupled to the EPM. Every time unit, the following sequence occurs:

- the *SIR model* sends its state variables values (S_0, I_0, R_0) to the *exp_plan_generator model*;
- the *exp_plan_generator model* checks whether the SIR state changed enough or not;
- if $\Delta S_0 < \Delta S_{max}$, the *exp_plan_generator model* doesn't do anything and the simulation goes on,
- if $\Delta S_0 > \Delta S_{max}$, the *exp_plan_generator model* builds the transfer Experimental Plan structure $EP_{transfer}$ as described in Table 2;

Table 2. Transfer Experimental Plan Structure ($EP_{transfer}$)

Variable	Value
Simulated duration	t_1
DEVS structure	IBM Model (see section 3.4.1.)
Dynamics	IBM Model (see section 3.4.1.)
Parameters	IBM Model (see section 3.4.1.)
Initial state	Randomly distributed individuals with state proportions (S_0, I_0, R_0)
Observation policy	each time unit observe: nb S, nb I, nb R
Nb replicates	$n_0 = 40$

- the *experimenter* model executes the transfer Experimental Plan and sends the corresponding results *ER* (see section 2.2.2.) to the *results_analyser* model;
- the *results_analyser* model uses the *ER* to compute the estimated β according to equations 5 and 4;
- the *results_analyser* model sends the β value to the *SIR* model, the *SIR* model changes its value of β and the simulation carries on.

**Figure 6.** Multiscale model: coupling the macroscopic model (SIR) and the scale transfer experimentation process model

We can give the complete specification of the experimentation process model structured as shown figure 6. As we said in the first part of the paper, the *experimenter* model is generic and already specified in section 2.2.2.. We just need to specify (i) the *exp_plan_generator* model and (ii) the *results_analyser* model as DEVS structures. Their complete

specifications are given in the appendices (see appendix A). In the two following sections we present our implementation of this model and then some simulations of the coupled model showing the values of the parameters we used.

3.4.4. Implementation

In this work, we have implemented all our models in a DEVS framework. To do so we model the SIR integrator as a Quantized State System [Kofman and Junco, 2001]. The experimenter model is a classical DEVS model. The complete implementation is realized within the Virtual Laboratory Environment (VLE) [Quesnel et al., 2009], a DEVS based set of tools and application programming interfaces for multi-modelling and simulation of complex systems. VLE proposes an XML representation for experimental plans which can be embedded in the state of a model. Moreover, issue of using embedded simulation for scale transfer were addressed within the VLE development team. Thus, the VLE environment enables embedded simulations, meaning that a DEVS model API contains functions allowing to run stand alone simulations using the VLE kernel simulation engine. The experimenter model is a DEVS atomic model with two particularities. Firstly, it contains in its state the experimental frame structure coded with the VLE XML library. Secondly, it is able to launch another VLE simulation and to get back the associated simulation results in an Experiment Results data structure that can be serialized to be sent on DEVS communication ports. This data structure can then be deserialized by the *results_analyser* model.

Staying in a "DEVS world" and using VLE facilitate the implementation of our models. One important feature of VLE is the possibility to distribute simulations defined in an experimental plan on several processors or computers. It is useful here, because several simulations of the experimented model can be performed simultaneously.

3.4.5. Simulations and results

We performed two virtual experiments with two different individual behaviours for the IBM (faster individuals and slower individuals) and observed the consequences at the macroscopic level.

We use the same set of parameters for both simulations except for the time step of individuals in the IBM that controls the speed of the individuals. Table 3 presents the values we used for all the parameters.

Simulation results are presented figure 7. We plotted the values of the S, I and R variables of the macroscopic simulation for both experiments (One chart for each experiment). The simulation's curves of the two graphics of figure 7 are classical for SIR models: At the beginning of the simulation, there are many susceptible individuals that can be infected

Table 3. Multiscale model's parameters

parameter	value ²
t_1	$20 t$
a	$80 \times 80 c$
infection probability	0.5
individual's time step	uniform distribution between $[0.8, 1] t$ (sim 1) between $[0.5, 0.7] t$ (sim 2)
mean infectious period	$150 t$
t_2	$750 t$
γ	$0.0033 t^{-1}$
d	$0.015625 \text{ individuals}.c^{-1}$
n_0	40
ΔS_{max}	$15 \text{ individuals}.c^{-1}$
T_0	$5 t$

by infectious individuals; Hence, the number of infected individuals rises at the same time as the number of susceptible individuals decreases; Then, at a point, the number of susceptible individuals to infect becomes too small, the recovering rate becomes higher than the force of infection and the number of infectious individuals start decreasing. The number of recovered individuals always rises while the infected individuals are recovering.

The interest of taking into account microscopic processes in similar models has already been discussed in the literature [Rahmandad and Sterman, 2008, Jacek and Wojciech, 1999, Keeling and Grenfell, 2000] and is not the question we want to address here. We rather want to illustrate that the dynamics of the coupled model can be parametrized at the microscopic scale.

The two simulations presented here (see results figure 7) illustrate this point. One can notice easily that if one changes the value of a parameter in the microscopic model (the speed of the individuals), the behaviour of the coupled model changes. In the simulation with slower individuals (top chart) the whole population has been infected after 25 time units (number of susceptible individuals reaches 0), and in the simulation with faster individuals (bottom chart), the whole population has been infected after only 15 time units.

Note that the parameter β does not appear in the table of parameters (table 3). Instead, we set the parameters of the IBM and of the EPM dedicated to scale-transfer (see section 3.4.3.). This application shows how multi-scale model can be specified using the EPM framework proposed in section 2.2..

²c and t are respectively surface and time arbitrary units ("c" corresponds to the surface of one cell of the IBM space).

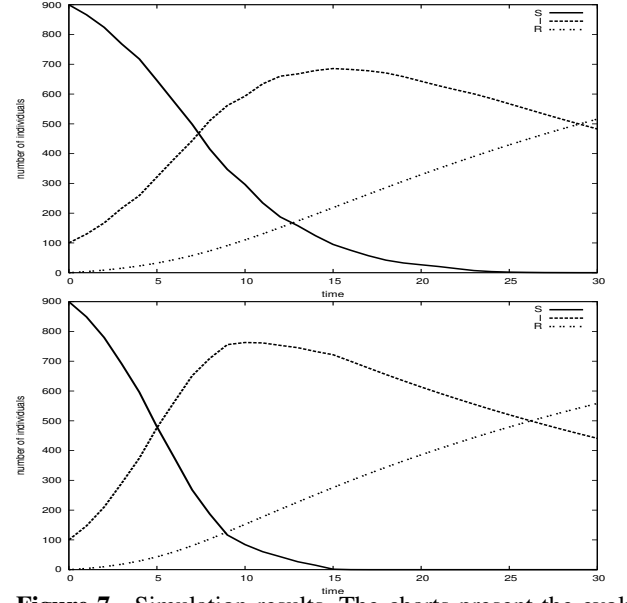


Figure 7. Simulation results. The charts present the evolution of the S, I and R, variables of the macroscopic model in the coupled simulations. The top chart shows the results of simulation 1 (with slower individuals) and the bottom chart shows the results of simulation 2 (with faster individuals).

4. DISCUSSION

4.1. EPM discussion, limits and perspectives

The experimental plan data structure specified in section 2.2.2. only contains the observation policy. In order to specify a list of input events per simulation and to specify how to compute in-line pre-treatment of the outputs during the simulations two adaptations can be made inside the experimented model itself. First, the list of input events must be specified as a combination of a DEVS model dynamics, initial state and parameters. Secondly, any in-line pre-treatment of model outputs can be specified in another DEVS model that would be observed itself following the observation policy. These models would correspond to dynamic components of the EF, such as *generator* and *acceptor*. In other words, at the stage of development of our framework, the experimental frame dynamics is to be specified as a DEVS model and coupled with the experimented model. The simulation of the experimental frame is performed by a DEVS simulator included within the embedded simulation. As a future work, we would like to use the specification of Frame systems defined in [Traore and Muzy, 2006]. Indeed, Traoré and Muzy proposed a framework to formalize the Experimental Frame at the highest level of specification. They also explain how to simulate the coupled model composed of the EF and the model, by introducing the concept of Experimenter to simulate the Experimental frame. This Experimenter is for the Frame what the simulator is for the model. It adheres to the

same generic message protocol. We did not implement it yet because it requires to slightly modify our simulation engine.

Note also that we only consider Experimental Frames that specify a single experiment. Further work would be necessary to bring the discussion up to higher levels of abstraction and include Experimental Frames specifying the "set of experiments of interest" within our framework. Likewise, we did not use concepts from experimental science such as workflows or knowledge domain ontologies [Maccagnan et al., 2010]. The link between such work (high level conceptual tools to describe experiments), and what we propose in this paper (an operational way to specify the simulation of the experimental process), will be done in a next step.

We must also discuss our definition of an experiment. Our definition (given in section 2.1.) differs from the usual definition in that we include the model specification within the experiment specification. CD++ models and Experimental Frame repository [Chreyh and Wainer, 2009] for example distinguishes clearly model specification from experiment specification. The same experiment can be done on two different models. Actually, in the Theory of Modelling and Simulation, a single experiment must be specified for both the source system and its model. We do include the model specification within the experiment specification because, in the EPM framework, we only consider virtual experiments. Thus, we do not lose any generality because the model must be specified in order to run the experiment. This definition fits with the operational way to specify virtual experiments in the modelling and simulation environment we use [Quesnel et al., 2009]. Of course, it fits also with our definition of iterative experimental design, illustrated figure 2. During such experimental design, experimental plans (as defined section 2.1.) are generated, then computed by the simulation environment and then analysed. In this framework, an experiment is simply the specification of a simulation, thus, it includes the model.

4.2. Multiscale modelling and simulation in epidemiology and ecology

We applied our framework to multiscale modelling and simulation in epidemiology and ecology. Let's discuss what such framework can bring to the subject. We are interested in the case where we want to represent macroscopic behaviour and need to represent microscopic mechanisms to do that. Such multiscale models are widely used in Physics and have been regrouped under the denomination of "Heterogeneous Multiscale Methods". These multiscale methods are used in the following way: *in cases when the macroscopic model is not explicitly available or invalid, the microscopic solver is used to supply the necessary data for the macroscopic model. Scale separation can be exploited to considerably reduce the complexity of the microscopic solver* [Weinan and Engquist, 2002].

Modelling in epidemiology (and in ecology) is a large scientific domain in which many specific aggregated and individual-based models have been developed [Keeling and Rohani, 2008]. Individual based models even use specific paradigms such as multi-agents systems [Bousquet and Page, 2004, Roche et al., 2008] or social networks models [Perkins et al., 2009]. These models are often specified with few mathematical formalism regarding the models used in Physical sciences. The most famous standard protocol for describing such individual-based models proposes to use several formalisms to describe a single model including diagrams, source code and mathematical equations (ODD, [Grimm et al., 2006]). In fact, even if some of these individual-based models can be represented with mathematical models, most of them cannot because they may involve complex decision rules [Treuil et al., 2008].

This lack of mathematical formalisation is not a problem as far as the model specification is rigorous enough. However, it induces the fact that transposition of methods like Heterogeneous Multiscale Methods from physical science to ecology or epidemiology is not straightforward. Therefore, we need to take a stand on modelling and simulation in ecology and epidemiology and observe as Aumann and Ginzburg that there is still need for methodological tools to be developed in these domains [Aumann, 2007, Ginzburg and Jensen, 2004]. Our application in multiscale modelling isn't validated yet, but it shows how we can use the EPM to formalize a Heterogeneous Multiscale Method. The use of the EPM enables an unambiguous description of the multiscale model.

5. CONCLUSION

The formal framework we have proposed for the experimentation process is based on the Experimental Frame concept, the recursive simulation technique and the DEVS formalism. We built a modular DEVS model of this experimentation process. As a DEVS model, the Experimentation Process Model (EPM) can be coupled with other DEVS models and become part of a wider model.

We have implemented the EPM in a DEVS modelling environment (VLE) in order to illustrate its concrete use and applicability. The first application was the specification of an original method for multiscale modelling and simulation. We gave an example in epidemiology. It gives insights on the potentials of the EPM by illustrating how results of some simulations can be used to feed another one. The method presented for multiscale modelling does not claim to enrich scale transfer theory, but proposes an original operational framework to specify multiscale models.

As we have said in the discussion, work remains to be done to have a complete representation of the experimentation process. However, the framework has been implemented and can be used for other applications, such as optimisation

or anticipation models for instance. Furthermore, as we have said in the introduction, its complete formalisation will insure reproducibility, an essential component of the scientific use of modelling and simulation. In agreement with Aumann [2007], we think that complex systems modelling, and more particularly ecological modelling (the application field of the authors), needs new methodologies and tools. We hope that the framework developed in this paper could be a useful contribution.

6. ACKNOWLEDGMENTS

We thank the three anonymous reviewers for their careful reading and valuable comments.

REFERENCES

- G. A. Aumann. A methodology for developing simulation models of complex systems. *Ecological Modelling*, 202: 385–396, 2007.
- F. J. Barros. Dynamic structure discrete event system specification: Formalism, abstract simulators and applications. *Winter Simulation*, 13(1):35–46, 1996.
- F. Bergero and E. Kofman. Powerdevs: a tool for hybrid system modeling and real-time simulation. *SIMULATION*, 87(1-2):113–132, 2011. doi: 10.1177/00375497110368029.
- B. Bonté, R. Duboz, G. Quesnel, and J.-P. Müller. Recursive simulation and experimental frame for multiscale simulation. In *SCSC'09: Summer Computer Simulation Conference*, July 13th-16th 2009.
- F. Bousquet and C. L. Page. Multi-agent simulations and ecosystem management: a review. *Ecological Modelling*, 176(3-4):313–332, 2004. ISSN 0304-3800. doi: DOI: 10.1016/j.ecolmodel.2004.01.011.
- M. Choisy, J.-F. Gueguan, and P. Rohani. *Encyclopedia of Infectious Diseases: Modern Methodologies*, chapter Mathematical Modeling of Infectious Diseases Dynamics. John Wiley and Sons Inc, 2007.
- R. Chreyh and G. Wainer. Cd++ repository: an internet based searchable database of devs models and their experimental frames. In *Proceedings of the 2009 Spring Simulation Multiconference*, SpringSim '09, pages 159:1–159:8, San Diego, CA, USA, 2009. Society for Computer Simulation International.
- R. Duboz, E. Ramat, and P. Preux. Scale transfer modelling: Using emergent computation for coupling an ordinary differential equation system with a reactive agent model. *Systems Analysis Modelling Simulation*, 43(6):793–814, june 2003.
- R. Duboz, D. Versmisse, G. Quesnel, A. Muzzy, and E. Ramat. Specification of dynamic structure discrete event multi-agent systems. In *Agent Directed Simulation (Spring Simulation Multiconference)*, Huntsville, Alabama, USA, April 2-6 2006.
- B. Durand, M. A. Dubois, P. Sabatier, D. Calavas, C. Ducrot, and A. V. de Wille. Multiscale modelling of scrapie epidemiology: 2. geographical level: hierarchical transfer of the herd model to the regional disease spread. *Ecological Modelling*, 179(4):515 – 531, 2004. ISSN 0304-3800. doi: DOI:10.1016/j.ecolmodel.2004.05.015.
- G. Gianni and P. Sara. Use of individual-based models for population parameters estimation. *Ecological Modelling*, 200:109–118, 2007.
- J. B. Gilmer and F. J. Sullivan. Issues in event analysis for recursive simulation. In M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. J. eds., editors, *Proceedings of the 2005 Winter Simulation Conference*, pages 1234–1241. The Society for Computer Simulation, july 2005.
- L. R. Ginzburg and C. X. J. Jensen. Rules of thumb for judging ecological theories. *Trends in Ecology & Evolution*, 19(3):121 – 126, 2004. ISSN 0169-5347. doi: DOI:10.1016/j.tree.2003.11.004.
- C. A. Goble, J. Bhagat, S. Alekseyevs, D. Cruickshank, D. Michaelides, D. Newman, M. Borkum, S. Bechhofer, M. Roos, P. Li, and D. De Roure. myExperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucleic Acids Research*, 38(suppl 2):W677–W682, 2010. doi: 10.1093/nar/gkq429.
- V. Grimm, U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jorgensen, W. M. Mooij, B. Muller, G. Pe'er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmann, N. Rüger, E. Strand, S. Souissi, R. A. Stillman, R. Vabo, U. Visser, and D. L. DeAngelis. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1-2):115 – 126, 2006. ISSN 0304-3800. doi: DOI: 10.1016/j.ecolmodel.2006.04.023.
- W. Jacek and J. Wojciech. Individual based modeling and parameter estimation for lotka-volterra system. *Mathematical Biosciences*, 157:23–36, 1999.
- M. J. Keeling and B. T. Grenfell. Individual-based perspectives on r0. *Journal of Theoretical Biology*, 203(1):51 – 61, 2000. ISSN 0022-5193. doi: DOI:10.1006/jtbi.1999.1064.
- M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in humans and animals*. Princeton University Press, 2008.

- E. Kofman and S. Junco. Quantized state systems. a DEVS approach for continuous systems simulation. In *Transactions of the Society for Computer Simulation International*, volume 18, pages 123–132. Society for Computer Simulation International San Diego, CA, USA, 2001.
- A. Maccagnan, M. Riva, E. Feltrin, B. Simionati, T. Vardanega, G. Valle, and N. Cannata. Combining ontologies and workflows to design formal protocols for biological laboratories. *Automated Experimentation*, 2(1):3, 2010. ISSN 1759-4499. doi: 10.1186/1759-4499-2-3.
- S. E. Perkins, F. Cagnacci, A. Stradiotto, D. Arnoldi, and P. J. Hudson. Comparison of social networks derived from ecological data: implication for inferring infectious disease dynamics. *Journal of Animal Ecology*, 78:1015–1022, 2009. doi: DOI:10.1111/j.1365-2656.2009.01557.x.
- N. Picard and A. Franc. Aggregation of an individual-based space-dependent model of forest dynamics into distribution-based and space independent models. *Ecological Modelling*, 145:69–84, 2001.
- G. Quesnel, R. Duboz, and E. Ramat. The Virtual Laboratory Environment – An operational framework for multi-modelling, simulation and analysis of complex dynamical systems. *Simulation Modelling Practice and Theory*, 17: 641–653, April 2009.
- H. Rahmandad and J. Sterman. Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models. *MANAGEMENT SCIENCE*, 54(5):998–1014, 2008. doi: 10.1287/mnsc.1070.0787.
- B. Roche, J.-F. Guegan, and F. Bousquet. Multi-agent systems in epidemiology: a first step for computational biology in the study of vector-borne disease transmission. *BMC Bioinformatics*, 9(1):435, 2008. ISSN 1471-2105. doi: 10.1186/1471-2105-9-435.
- P. Sabatier, B. Durand, M. Dubois, C. Ducrot, D. Calavas, and A. V. de Wielle. Multiscale modelling of scrapie epidemiology: 1. herd level: a discrete model of disease transmission in a sheep flock. *Ecological Modelling*, 180(2-3): 233 – 252, 2004. ISSN 0304-3800. doi: DOI:10.1016/j.ecolmodel.2004.05.012.
- M. K. Traore and A. Muzy. Capturing the dual relationship between simulation models and their context. *Simulation Modelling Practice and Theory*, 14:126–142, 2006.
- J.-P. Treuil, A. Drogoul, and J.-D. Zucker. *Modelisation et simulation a base d'agents - Exemples commentes, outils informatiques et questions theoriques*. 2008.
- G. A. Wainer and N. Giambiasi. Application of the Cell-DEVS paradigm for cell spaces modelling and simulation. In *Simulation*, volume 76, pages 22–39, 2001.
- E. Weinan and B. Engquist. The heterogeneous multi-scale methods. *Comm. Math. Sci*, 1:87–132, 2002.
- B. P. Zeigler, T. G. Kim, and H. Praehofer. *Theory of modeling and simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems*. Academic Press, 2000.

Appendices

A SPECIFICATION OF EPM SUB-MODELS FOR MULTI-SCALE MODELLING

i. *exp_plan_generator* model:

- S , the set of states is the following:

$$S = \{S_{old}, SIR, phase | phase \in WAIT, GEN\}, SIR \in \mathbb{R}^3 \text{ and } S_{old} \in \mathbb{R}$$

- X , the set of input events is \mathbb{R}^3 .
- Y , the set of output events is the set of possible experimental plans (EP).
- δ_{ext} , the external transition function is the following:

$$\delta_{ext} : X \times S \rightarrow S$$

$$\{(S_0, I_0, R_0), (S_{old}, SIR, WAIT) | S_{old} - S_0 \leq \Delta_{max}\} \mapsto (S_{old}, SIR, WAIT)$$

$$\{(S_0, I_0, R_0), (S_{old}, SIR, WAIT) | S_{old} - S_0 > \Delta_{max}\} \mapsto (S_0, (S_0, I_0, R_0), GEN)$$

- ta , the time advance function is the following:

$$ta : S \rightarrow \mathbb{R}_0^+ \cup \infty$$

$$(S_{old}, GEN) \mapsto 0$$

$$(S_{old}, WAIT) \mapsto \infty$$

- δ_{int} , the internal transition function does nothing:

$$\delta_{int} : S \rightarrow S$$

$$\forall s \in S : s \mapsto s$$

- λ , the output function is:

$$\lambda : S \rightarrow Y$$

$$(S_{old}, SIR, GEN) \mapsto EP$$

Where EP is built as shown table 2

- δ_{con} , the confluent transition function is:

$$\delta_{con}(s, x) = \delta_{int}(\delta_{ext}(s, ta(s), x))$$

ii. *results_analyser* model:

- S , the set of states is the following:

$$S = \{\beta | \beta \in \mathbb{R}_*^+\}$$

- X , the set of input events is the set of possible Experiment Results data structures ER containing $n_0 \times t_1$ triplets $(s, i, r) \in \mathbb{R}^3$.
- Y , the set of output events is \mathbb{R}_*^+ .
- δ_{ext} , the external transition function is the following:

$$\delta_{ext} : X \times S \rightarrow S$$

$$(ER), (\beta) \mapsto (\beta_{new})$$

Where β_{new} is computed with equation 5.

- ta , the time advance function is the following:

$$ta : S \rightarrow \mathbb{R}_0^+ \cup \infty$$

$$\forall s \in S : s \mapsto 0$$

- δ_{int} , the internal transition function does nothing:

$$\delta_{int} : S \rightarrow S$$

$$\forall s \in S : s \mapsto s$$

- λ , the output function is:

$$\lambda : S \rightarrow Y$$

$$\beta \mapsto \beta$$

- δ_{con} , the confluent transition function is:

$$\delta_{con}(s, x) = \delta_{int}(\delta_{ext}(s, ta(s), x))$$

Annexe G

Vers une spécification des modèles de simulation de systèmes complexes

Vers une spécification des modèles de simulation de systèmes complexes

Raphaël Duboz^{1,2} — Bruno Bonté¹ — Gauthier Quesnel³

¹ CIRAD, UPR AGIRs 22, Campus International de Baillarguet,
34398 Montpellier Cedex 5, France

raphael.duboz@cirad.fr

² Asian Institute of Technology, Computer Science and Information Managment
Laboratory, P.O. Box 4, Klong Luang, Pathumthani 12120, Thailand

³ INRA, UR875 Biométrie et Intelligence Artificielle,
F-31326 Castanet-Tolosan, France

gauthier.quesnel@toulouse.inra.fr

RÉSUMÉ. *L'adoption d'un protocole partagé pour communiquer les modèles de simulation de systèmes complexes de façon rigoureuse est actuellement nécessaire. Ce constat est soutenu par la demande croissante en modélisation, non seulement pour comprendre la réalité qui nous entoure, mais également pour faire de l'aide à la décision. La spécification rigoureuse des ces modèles de simulation est essentielle car elle permet leur communication, leur reproduction et donc leur vérification. Dans ce papier, nous proposons de réunir les propositions de deux communautés en modélisation et simulation. La première est la communauté des chercheurs en modélisation individus centrés, ou modélisation à base d'agents. Cette communauté propose un protocole, défini par V. Grimm et al. en 2006, revu en 2010. La seconde est la proposition issue de la communauté des chercheurs en simulation au sens large, qui existe autour de la théorie initiée par B. Zeigler en 1976. Nous faisons le lien entre ces deux communautés à partir d'un article de G. Aumann de 2007 qui introduit la théorie de Zeigler en écologie. Ici, nous proposons et décrivons le cadre général d'une méthode de spécification des modèles de simulation qui permet la reproduction et la vérification, et donc augmente la confiance autour des modèles de simulation de systèmes complexes. Nous proposons l'utilisation d'un formalisme systémique pour la spécification de ce protocole. Néanmoins les aspects techniques et purement formels de la spécification, même s'ils sont fondamentaux, ne sont pas abordés dans ce papier car ils ne*

sont pas nécessaires à sa compréhension. Ils devront néanmoins être présentés prochainement dans leur totalité pour que le protocole proposé ici soit effectif.

ABSTRACT. *TODO*

MOTS-CLÉS : *Système, Formalisme, Communication, Protocole, cadre expérimental, Multiagents, ODD.*

KEYWORDS: *TODO*

1. Introduction

La communication rigoureuse d'un protocole expérimental est au cœur de l'activité scientifique. Elle permet la reproduction sans ambiguïté de l'expérience, et donc la vérification des résultats par les pairs. Dans les sciences expérimentales, la description des protocoles et des appareils de mesures peut être longue et laborieuse. Elle consiste en une documentation technique qui décrit les appareils, les phénomènes physico-chimiques en jeu, les calculs associés pour restituer l'information etc. Ces appareils font l'objet de procédures de normalisation pour garantir la fiabilité et la reproductibilité des résultats. Il devrait en être de même pour les modèles de simulation, c'est-à-dire les modèles qui nécessitent de passer par une résolution numérique, ou disons plus généralement par un calcul sur ordinateur, pour en explorer le comportement, la dynamique. Actuellement, les simulations sont considérées comme des expériences virtuelles, ou expériences de seconds ordres [Var06]. Elles permettent de résoudre des équations autrement insolubles ou d'exécuter des algorithmes qui simulent le comportement d'un système donné. La simulation prend ainsi une place de plus en plus importante dans l'étude des systèmes complexes, en tant qu'objet de substitution au réel. Ainsi, faire des expériences à l'aide d'un modèle de simulation devient une pratique courante pour tester des hypothèses sur les dynamiques et les structures (*c.-à-d.* les composants et leurs relations) des systèmes complexes. Néanmoins, les modèles de simulation peuvent être compliqués, c'est-à-dire longs et difficiles à décrire, et leurs comportements aussi complexes que les phénomènes qu'ils sont censés représenter. Ces deux points sont à l'origine de la réticence de certains scientifiques à utiliser ce type de modèle. Ces scientifiques mettent en doute la confiance que l'on peut avoir dans les résultats obtenus par simulation et donc dans les conclusions que l'on peut en tirer (voir à ce sujet l'introduction de l'article de Graig A. Aumann [Aum07], pour un argumentaire développé pour l'écologie).

Il ne s'agit pas pour nous de discuter ici de l'intérêt d'avoir ou non des modèles de simulation très détaillés. Il est question de faire le point sur la spécification des modèles de simulation en établissant un lien entre deux approches différentes. La première approche est un protocole de communication des modèles individus centrés (IBM pour *Indi-*

vidual Based Model) ou modèles de simulation multiagents (ABS pour *Agents Based Simulation model*) développée par V. Grimm et al. en 2006 [VUF⁺06], revue et améliorée en 2010 [GUD⁺10]. La seconde est celle développée depuis les années 70 par B.P. Zeigler et ses évolutions récentes [Zei76, ZKP00, Wai10] dans le cadre de la Théorie de la Modélisation et de la Simulation (TMS).

La problématique de la communication et de la vérification est centrale et bien identifiée, à la fois par la communauté de la TMS ainsi que par la communauté des concepteurs et utilisateurs d'IBM et d'ABS. La première communauté s'intéresse surtout, mais pas seulement à des simulations orientées vers l'optimisation et la conception de systèmes industriels ou plus généralement artificiels. Dans ce cas, le système est suffisamment bien connu pour pouvoir en reproduire le comportement. La deuxième communauté utilise principalement les simulations pour évaluer des représentations possibles de systèmes afin d'augmenter la connaissance sur les systèmes en question. Dans les deux cas, une spécification rigoureuse du modèle est nécessaire. Il s'avère que la communauté TMS propose un cadre général qui peut bénéficier à la communauté ABS, qui en retour pose des problèmes nouveaux à la TMS dans le domaine de la complexité [Aum07]. C'est ce qui a motivé cet article.

Dans une première partie, nous ferons un point rapide sur la spécification des ABS en général. Nous rappellerons ensuite certains des éléments du cadre théorique de la modélisation des systèmes. Nous ferons enfin une proposition méthodologique pour la spécification des modèles de simulation de systèmes complexes et nous la discuterons.

2. Spécification des simulations à base d'agents

Le cas des ABS est intéressant car le paradigme de modélisation des Systèmes Multi-Agents (SMA) dont ils sont issus, n'est pas compris, et par conséquent formalisés, de façon univoque (le livre récent de J.P. Treuil et al. illustre bien ce fait par les différents exemples de simulation à base d'agents qu'il présente [TDZ08]). La souplesse des définitions du paradigme agent n'est pas une tare, c'est un avantage qui permet à différentes disciplines aux définitions hétérogènes de se rencontrer sur un terrain sémantique commun. Néanmoins, quand vient le

moment des choix précis de modélisation, c'est-à-dire des formalismes et des algorithmes à implémenter pour simuler le modèle, les ambiguïtés doivent être levées, au moins entre le modèle et le simulateur, nous y reviendrons plus tard.

Prenons le cas emblématique de NetLogo, un logiciel de modélisation et simulation très utilisé dans le domaine des ABS [Wil99]. NetLogo propose un langage dit « de modélisation ». Ce langage consiste en des primitives de très haut niveau comme « *create-turtles* » pour créer des agents ou « *move-to* » pour déplacer un agent. La syntaxe du langage NetLogo permet d'écrire des algorithmes qui simulent le comportement des agents, des groupes d'agents et de leurs environnements. Au delà des aspects syntaxiques, ce type de langage de programmation agent est conditionné par les hypothèses sur ce qu'est un agent, son environnement, une interaction, une communication etc. Nous ne pouvons donc pas l'utiliser pour reproduire et vérifier des modèles de simulation à base d'agents conçu pour un autre logiciel, qui proposera une implémentation différente, même légèrement, des ABS. Il nous faudra ré-implémenter complètement le modèle, mais « *à l'aveugle* », car les hypothèses de conception ne sont pas clairement formulées. Des tentatives de mise au point de langages de modélisation agent existent. Par exemple en utilisant UML et en générant le code [PTW07], ou les langages métiers [Hah08], ou même des formalisation systématique [DRQ04], mais il semble qu'ils soient très peu utilisés de par leur manque d'expressivité et surtout de généricité [SDS10]. Pourtant chaque simulation d'agents n'est pas seulement un cas particulier, et un protocole à été proposé pour leur spécification.

2.1. Ambiguïtés du protocole « Overview Desing concepts Details » (ODD)

Une méthodologie de description des ABS à été proposée par V. Grimm et al. en 2006 [VUF⁺06], revue et améliorée en 2010 [GUD⁺10], il s'agit du protocole ODD pour « *Overview Desing concepts Details* ». Ce protocole améliore la communication des ABS, mais des imprécisions et des ambiguïtés demeurent qui empêchent, à notre sens, l'utilisation d'ODD pour la communication et la vérification

des modèles de simulation, ce qui était pourtant l'objectif affiché par ce protocole dès 2006, lorsque les auteurs disaient qu'un des objectifs à long terme d'ODD était d'arriver à un niveau de description suffisamment précis pour que la description fournie puisse être « convertie directement en un simulateur sans possibilité d'erreurs de programmation ». Les auteurs ajoutent qu'une fois cet exécutable généré, un environnement d'expérimentation de modèle doit être mis à disposition pour permette de générer, d'observer, d'interpréter, etc. des expériences. Ce dernier point implique que cet environnement d'expérimentation soit lui-même décrit sans ambiguïté. Néanmoins, l'objectif de 2006 ne semble toujours pas atteint, et nous donnons ici des exemples d'ambiguïtés relevées dans le protocole ODD (se référer à l'article [GUD⁺10] pour les détails du protocole).

2.2. Confusion entre expérience et modèle

La différence entre expérience et modèle est bien sûr fondamentale et nous le discuterons plus loin. Quoi qu'il en soit, les valeurs des paramètres qui font partie de l'expérience faite sur un modèle, ainsi que leur définition, doivent à notre sens apparaître à un seul endroit pour éviter les confusions. Hors dans le protocole ODD, les définitions et les valeurs des paramètres peuvent apparaître à plusieurs endroits : section 7 *Submodels*, section 5 *Initialization* ou dans la définition des valeurs initiales des variables d'états. Ces variables d'états (également appelées attributs dans le protocole ODD) peuvent rester constantes lors de l'exécution du modèle. Si c'est le choix du modélisateur et non une conséquence du calcul, alors une variable qui reste constante n'est plus une variable, par définition, c'est un paramètre. En plus de possibilités multiples de déclaration des paramètres, la distinction n'est pas faite clairement entre variables et paramètres dans le protocole ODD.

2.3. Définition de la notion de variable d'état

Le fait de définir une variable d'état comme ne pouvant pas être calculée à partir d'autres variables d'états est également très restrictif. Si l'on prend l'exemple donné pour décrire le protocole, la distance entre

un agent localisé par x et y sur une grille et un service quelconque localisé sur une autre case de la même grille, ne devrait pas être une variable d'état car elle est calculable à partir d'autres variables d'états du modèle. Donc, si la position de l'agent est calculée en fonction de sa vitesse et de son comportement, alors elle ne plus n'est pas une variable d'état, pour les mêmes raisons. Nous pouvons alors nous poser la question légitime de savoir qu'est-ce qui est « variable d'état » dans un modèle où presque toutes les valeurs dynamiques sont calculées en fonction d'autres valeurs dynamiques ? De plus, cette définition interdit de considérer une valeur agrégée (la somme des agents présents à un instant t par exemple) comme une variable d'état, alors que cette valeur pourrait très bien servir à un agent collectif pour entreprendre une action par exemple. Cette valeur qualifierait bien l'état du collectif à un instant t mais ne serait pas une variable d'état décrite par le protocole ODD.

2.4. Spécification des modèles

Dans le protocole ODD, la description des « *sub-models* » peut prendre des formes verbales, mathématiques ou algorithmiques par l'ajout du code informatique complet en annexe du papier. Encore une fois, cette totale liberté d'expression de l'ensemble des modèles composants un ABS ne peut pas être considérée comme une avancée vers une meilleure communication des modèles, car aucune standardisation ou normalisation n'est possible ici du fait de la multiplicité des langages et des simulateurs, qui sont de plus amenés à évoluer, ce qui rend ce type de spécification non pérenne.

Nous pensons qu'il y a parfois une confusion entre modélisation objet et modélisation des systèmes dynamiques dans le protocole ODD (héritage, versus niveau de spécification et base de temps, voir section 3.1). Dans la version révisée du protocole [GUD⁺10], l'excessive proximité entre la programmation objet et ODD a été identifiée. Les auteurs reviennent notamment sur leur conseil initial d'utiliser les diagrammes de classes UML pour décrire graphiquement la structure des modèles. Ils persistent cependant à décrire les entités de leurs modèles comme un ensemble d'attributs et de méthodes ou de fonctions, c'est-

à-dire comme une classe ou son instance en langage objet. Les auteurs expliquent que le lien entre les entités et les processus associés doivent être décrits dans la partie « ordonnancement » de la spécification du modèle global. Néanmoins, la représentation du temps, et donc de la dynamique des interactions entre agents ou sous-modèles, reste problématique dans cette proposition.

3. Spécification dans la théorie de la modélisation et de la simulation

3.1. Les entités de base

Nous rappelons ici les définitions qui sont utilisées dans la suite de cet article. Nous ne donnerons pas ici les définitions formelles, elles sont accessibles avec les références citées dans cette partie. Ces définitions ne sont pas indispensables pour une présentation générale.

Une notion importante est celle de « trajectoire d'états » pour les systèmes dynamiques. Les simulations génèrent des trajectoires d'états, c'est-à-dire des suites de valeurs ou de symboles indexées par le temps. Elles sont la trace tangible du comportement du modèle. Ce sont ces trajectoires d'états qui servent de support aux discussions autour de la vérification du modèle, de sa validation, et de son adéquation avec les questions posées.

Une notion également importante est celle de « niveau de spécification » [Kli85]). Cette notion met en relation le niveau de connaissance que l'on a d'un système donné avec sa spécification formelle. Plus le niveau est bas, moins le système est connu dans ces détails. Par exemple au premier niveau, le niveau 0, on dispose juste de données en entrée et en sortie du système, indexées par le temps, c'est le niveau de l'observation du comportement du système.

Les quatre entités de base en Théorie de la Modélisation et de la Simulation (TMS) sont [ZKP00] :

- *Le système source* : c'est le système étudié, ou plus généralement le système d'intérêt. Il peut être réel ou virtuel. Il est vu comme une source ou une base de données, de trajectoires d'états. Il est spécifié au

niveau 0.

– *Le modèle* : à un niveau de spécification donnée au sens de [Kli85], un modèle est un ensemble de règles, d'équations, de contraintes qui spécifie des relations dynamiques entre des entrées et des sorties. Il s'agit de la description d'un mécanisme de génération de trajectoires de sortie en fonction de trajectoires d'entrée. Cette définition à l'avantage d'avoir une base mathématique dans la théorie des systèmes qui permet la communication non ambiguë des modèles. Elle permet également de considérer un modèle à l'intérieur d'un modèle plus grand, en utilisant des mécanismes de composition par couplage de modèles [ZKP00].

– *Le simulateur* : c'est l'entité qui génère le comportement d'entrée-sortie décrit par le modèle. Il peut s'agir d'un calculateur, d'un mécanisme physique quelconque ou encore du cerveau humain.

– *Le cadre expérimental*.

Cette dernière entité est importante pour la spécification des modèles de simulation car elle englobe le modèle, le simulateur et le système source dans un cadre d'usage (figure 1). Le cadre expérimental permet de mettre en relation le système source et le modèle. Nous le détaillons dans ce qui suit.

3.2. Spécification des cadres expérimentaux

À partir des objectifs de modélisation, il est possible de définir un cadre expérimental. C'est précisément cette transformation des objectifs en cadre expérimental qui guide la conception du modèle et les expériences effectuées. Le cadre expérimental définit des observables (des trajectoires d'états observées) pour le systèmes source et le simulateur, qui devront être mis en relation pour statuer sur la validité du modèle, au regard des objectifs de modélisation. Les objectifs peuvent être de deux catégories différentes. Soit il s'agit de simuler un système existant ou virtuel dont on connaît le fonctionnement et la dynamique et dont on veut optimiser certaines parties, tester des configuration différentes, améliorer le fonctionnement etc. Les objectifs sont alors ceux de l'ingénierie des systèmes. Soit il s'agit, comme c'est le cas dans l'étude des systèmes complexes, de simuler un comportement pour inférer des règles de fonctionnement, découvrir, comprendre un mécanisme, le rôle

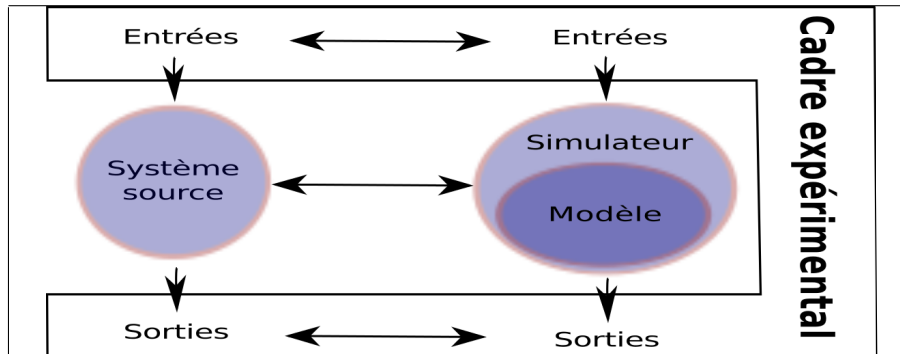


Figure 1 – Cadre général de la modélisation et de la simulation selon Zeigler [ZKP00] et redessiné par [Aum07]. Les flèches horizontales représentent les relations d'équivalence et d'approximation qui existent entre les entités du cadre. Les flèches verticales représentent un flux de données. La superposition du simulateur et du modèle illustre une relation d'équivalence particulière, un morphisme, que nous discutons dans le texte.

des interactions entre composants du système etc. C'est cette deuxième catégorie d'objectifs qui nous intéresse ici. Nous parlerons dans ce cas plutôt de « question posée » que d'« objectif ».

Pour relier le système source au modèle, la TMS associe la question posée à un cadre expérimental, c'est-à-dire à un contexte dans lequel le comportement du système source nous intéresse. Ce cadre expérimental peut être décrit formellement à différents degrés de spécification [Kli85, ZKP00]. Au niveau le moins précis, le cadre définit uniquement les domaines de définitions des entrées et des sorties du système source (et donc du modèle associé). Au niveau le plus précis, il s'agit d'un ensemble de composants qui interagissent avec le modèle et qui nous permettent à la fois de simuler le contexte d'intérêt, mais aussi d'observer et d'analyser les trajectoires des variables d'états qui nous intéressent. De manière générale, le cadre expérimental doit décrire rigoureusement l'ensemble des expériences possibles qui peuvent être réalisées sur le modèle dans le cadre de la question posée. Une telle approche permet de relier rigoureusement la ou les questions posées,

à l'ensemble des expériences. Traoré et Muzy proposent une description formelle de cadres expérimentaux à différents niveaux de spécifications [TM06]. La formalisation de ces cadres nous permet d'appliquer un même cadre expérimental à plusieurs modèles et réciproquement de tester un même modèle dans des cadres expérimentaux différents. Le premier cas nous permet par exemple de trouver le meilleur modèle ou un agencement préférentiel pour notre système si notre objectif est de concevoir un nouveau système. Le second cas nous permet quant à lui de réutiliser un même modèle pour une question légèrement différente qui considérera le système étudié dans un contexte différent ou sous un point de vue différent. Dans leurs travaux, Traoré et Muzy [TM06] partent d'un objectif précis pour arriver à un cadre expérimental entièrement spécifié dans lequel ils peuvent tester des modèles différents pour choisir le plus adapté. Dans leur exemple, ils cherchent un modèle de feu de forêt qui leur permette de prendre des décisions de lutte contre le feu en temps réel. Le cadre expérimental qu'ils définissent, simulé en même temps que le modèle, leur permet de tester différents scénarios d'occurrence des départs de feux tout en contrôlant les deux critères qui les intéressent : le temps d'exécution des simulations et l'erreur entre les résultats de simulation du modèle testé et ceux d'un modèle de référence. Le choix d'un modèle ou d'un autre ne va pas se faire au seul regard de sa faculté à imiter le système, mais sur un compromis entre les deux critères définis dans le cadre expérimental.

Spécifier une expérience consiste à définir un état initial et un ensemble de valeurs des paramètres du cadre expérimental. Par exemple, dans le cas de modèles de feux de forêts, la série temporelle des départs de feux est un paramètre du cadre expérimental. Il s'agit bien de la spécification d'une expérience qui aurait pu être réalisée sur le système. L'expérience consiste à déclencher des feux à certains moments à certains endroits. Le système correspond à la forêt. Les notions de cadre expérimental et d'expérience ainsi définies permettent de distinguer les paramètres de l'expérience de ceux propres au modèle.

Il est important de faire la différence entre un plan d'expériences et un cadre expérimental. Un plan d'expériences spécifie l'ensemble des valeurs prises par les paramètres et l'état initial des variables pour chaque simulation, ceci à la fois pour le modèle et pour le cadre ex-

périmental. Il va consister en une liste d'instances « cadre expérimental/Modèle », à tester. Un plan d'expérience est défini pour faire une analyse de sensibilité par exemple, ou optimiser un modèle au regard de certains critères définis a priori. Comme le montre le figure 1, le cadre expérimental englobe plus de choses que la spécification des valeurs de paramètres et de variables. Le plan d'expérience contient en fait une réalisation des trajectoires d'états en entrée du modèle, le cadre expérimental définit quand à lui l'ensemble des trajectoires d'états admissibles en entrée.

En TMS, la distinction claire, explicite du modèle et de son contexte d'utilisation trouve son origine dans les sciences de l'ingénieur et leurs applications. Les objectifs de cette communauté appartiennent à la première catégorie citée plus haut (tester ou évaluer des protocoles ou des agencements de systèmes devant être conçus ou utilisés par l'industrie). Les ingénieurs peuvent donc décrire clairement les attentes qu'ils ont sur leurs systèmes, tout comme ils peuvent limiter leur domaine d'intérêt à certaines conditions d'utilisation. De plus, l'activité de modélisation et simulation en sciences de l'ingénierie (écriture du simulateur incluse) est vue comme la production d'un bien (le modèle et son simulateur) qui doit être réutilisable et qui doit donc être livrée avec ses spécifications, c'est-à-dire avec son cadre expérimental général couvrant l'ensemble des cadres expérimentaux plus restreints pouvant être dérivés selon la question posée par l'utilisateur du modèle.

3.3. Distinction entre modèle et simulateur

La distinction entre modèle et simulateur est centrale en TMS. Selon la définition donnée section 3.1, un modèle correspond à une spécification. La transformation de cette spécification en une suite d'instructions exécutables par un ordinateur correspond à l'implémentation du modèle. Selon la définition, le code informatique d'exécution d'un modèle peut être considéré comme équivalent à sa spécification. Néanmoins, l'information présente dans ce code est spécifique à un langage et bien plus détaillée que l'information nécessaire et suffisante à la communication du modèle. Dans les faits, les modèles sont donc spécifiés dans des formalismes, mathématiques ou non, qui contiennent en théo-

rie l'ensemble de l'information nécessaire pour les décrire. L'étude de la relation d'équivalence entre le modèle et le simulateur est appelée « vérification » en TMS. Elle est, nous l'aurons compris, différente de la notion de « validation » car elle n'étudie pas la relation avec le système source. Cette activité de vérification consiste donc à s'assurer que le simulateur génère bien le comportement spécifié par le modèle. En effet, ce comportement dépend du langage de programmation, du compilateur ou même de l'architecture matérielle (par exemple, en fonction des optimisations du calcul des nombres flottants, des compilateurs, des systèmes d'exploitation, on aura pas exactement les mêmes résultats sur des architectures 32 ou 64 bits). Il dépend aussi de la phase d'implémentation et de ses erreurs potentielles. La vérification formelle du comportement des modèles de systèmes dynamiques est impossible dans beaucoup de cas, notamment lorsque le comportement est stochastique et que l'ensemble des états atteignables par le modèle est infini. Néanmoins les communautés de la TMS et du génie logiciel propose des solutions dans ce domaine de recherche toujours d'actualité [BCK09, LW05].

La vérification repose sur le concept d'homomorphisme qui définit si des systèmes sont équivalents en terme de transitions d'états [ZKP00]. Ce concept s'applique de façon général entre des systèmes quelconques. Prenons par exemple deux systèmes S et S' . Le système S peut être le système source et S' un système qui simule de façon simplifiée le comportement de S . Nous pourrions dire que S et S' sont homomorphes si pour toute transition d'états dans S , il existe une seule ou une suite de transitions d'états dans S' qui commence et finit par les états correspondants dans S . Si le système S est un modèle, S' les algorithmes du simulateur correspondant au modèle et S'' un calculateur qui exécute les instructions et que nous considérons un changement de l'état A vers l'état B , nous avons pour les trois systèmes considérés :

1) modèle

$$S : A \rightarrow B$$

2) algorithmes

$$S' : A' \rightarrow A'_1 \rightarrow A'_2 \rightarrow \dots \rightarrow A'_m \rightarrow B'$$

3) simulateur

$$S'' : A'' \rightarrow A''_1 \rightarrow A''_{11} \rightarrow A''_{12} \rightarrow \dots A''_{1n} \rightarrow A''_2 \rightarrow \dots \rightarrow A''_m \rightarrow B''$$

Ou, pour paraphraser les trois points précédents, il existe un ensemble de micro-transitions d'état dans un simulateur permettant de générer la transition décrite au niveau du modèle. Nous dirons qu'un simulateur est vérifié s'il y a un homomorphisme entre S et S'' . Nous voyons bien sur cet exemple que cela implique également un homomorphisme avec le système défini par les algorithmes de simulation. L'homomorphisme entre S' et S'' est assuré par le compilateur ou l'interpréteur du code informatique correspondant aux algorithmes de simulation. Même si des erreurs peuvent exister à ce niveau elle sont principalement d'ordre matérielles et nous n'en parlerons pas ici. La correspondance entre S et S' est plus compliquée qu'entre S' et S'' . Si cette correspondance est calculable, alors cela revient au cas de la correspondance entre S' et S'' . C'est le cas lorsque l'on dispose d'un langage de modélisation directement compilable ou interprétable (il en existe de très nombreux développés pour des formalismes particuliers comme les *states charts* d'UML, les réseaux de Petri, les automates à états etc). Néanmoins, ces langages imposent un type de modélisation car ils sont basés sur un formalisme particulier. En TMS, la distinction du modèle et du simulateur a permis de développer des outils qui permettent l'intégration de modèles hétérogènes [DRQ04], c'est-à-dire des compositions de modèles pluriformalisés augmentant le potentiel de description d'un système source par un modèle. C'est ce type de multimodèles qui sont difficiles à spécifier en raison de l'articulation compliquée des modèles composants. Là encore, la TMS propose une solution intéressante.

3.4. *Propriété de fermeture*

En TMS, le couplage de plusieurs modèles est vu comme l'union des états et des fonctions des modèles couplés. Le résultat de cette union est un nouveau modèle qui a un comportement équivalent au modèle couplé dont il est issu. L'équivalence du modèle atomique et du modèle couplé est appelée propriété de fermeture sous couplage [ZKP00]. Cette propriété peut être démontrée dans un formalisme donné et a des implications majeures en terme de modélisation. Par exemple, les entrées/sorties d'un modèle couplé et d'un modèle élémentaire sont spécifiées de la même manière *c.-à-d.* les couplages en entrée et sortie du modèle sont formalisés indépendamment de la structure interne des mo-

dèles et reste identique quelque soit le niveau de couplage. Ceci permet de définir des interfaces génériques pour le couplage de modèles, et donc de permettre la composition et la décomposition des modèles sans réécriture.

La propriété de fermeture garantit également que le choix de la structure du couplage n'ait pas d'influence sur le comportement du modèle, *c.-à-d.* sur les trajectoires d'états générées. Ceci est très important, car le choix de la granularité en terme de modèles élémentaires n'est pas unique. Cette absence d'influence de la structure de composition du modèle sur les trajectoires des états relègue toute la responsabilité des trajectoires d'états observées au niveau des choix de modélisation.

3.5. *Théorie des hiérarchies*

En TMS, la possibilité de composition hiérarchique de modèles fait naturellement penser à la théorie des hiérarchies [Bra73]. Dans cette théorie, une entité à un niveau hiérarchique donné (ou niveau d'organisation) contient en substance (physiquement, structurellement) les entités de hiérarchies inférieures et fait partie d'un niveau hiérarchique supérieur [Fei54]. En se basant sur cette théorie, G. Aumann identifie au minimum trois niveaux hiérarchiques pour tout modèle. Il les appelle les niveaux focaux [Aum07]. Il définit le niveau focal N comme le niveau d'intérêt dans le modèle, par exemple le niveau individuel dans un ABS ou le niveau population dans un modèle plus agrégé. C'est le niveau où sont modélisés les mécanismes d'intérêts, les interactions entre variables et composants du modèle. Le niveau focal $N - 1$ englobe les composants du modèle qui reproduisent un comportement existant à un niveau d'organisation inférieur. C'est le niveau de fermeture inférieur du modèle. Le niveau focal $N + 1$ correspond au comportement du système modélisé dans son ensemble.

Il est important de remarquer que les hiérarchies ainsi définies ne correspondent pas aux niveaux de spécification au sens de G. Klir [Kli85]. Il est par exemple possible d'avoir une très grande connaissance (niveau de spécification élevé) d'un mécanisme interne à un individu, au niveau focal $N - 1$. G. Aumann propose d'associer trois cadres expérimentaux aux trois niveaux focaux. Ce faisant, les trajectoires d'états en entrée et

sortie du modèle sont également associées aux trois niveaux focaux du modèle. De notre point de vue, G. Aumann clarifie ainsi l'étude de la dynamique des modèles de simulation de systèmes complexes en mettant en relation la théorie des hiérarchies et les modèles de simulation. Dans ce cadre, nous pouvons distinguer les trajectoires d'états qui sont le résultat de calculs qui reproduisent le comportement d'un sous-système, sans pouvoir explicatif au niveau $N - 1$ par exemple, des trajectoires d'états qui peuvent être qualifiées d'émergentes, car résultant des interactions des composants au niveau N et observables au niveau $N + 1$.

4. Proposition

4.1. *Présentation générale*

En adoptant une vision systémique des modèles de simulation, nous défendons une approche de construction et de spécification par composition de modèles. Chaque modèle étant conçu dans un cadre expérimental donné (pour répondre à des objectifs), la composition des cadres expérimentaux doit accompagner la composition des modèles [Aum07]. La figure 2 illustre les quatre bases de notre proposition¹. Il s'agit de séparer les spécifications des modèles de celles des cadres expérimentaux, de celles des structures (qui correspondent à la composition des modèles et donc des cadres associés). La base des expériences contient les plans d'expériences, instance des cadres expérimentaux. La structure d'un modèle composé est vue comme un arbre où chaque feuille correspond à un modèle indécomposable et chaque nœud à une composition. A chaque feuille et chaque nœud correspond un cadre expérimental.

L'avantage d'une telle représentation est qu'à partir d'une base de modèles donnée, toute nouvelle composition ne nécessite que la définition d'un nouveau cadre expérimental qui se trouve être la composition des cadres associés aux modèles composants. Les ABS sont un archétype de modèle de simulation de systèmes complexes. Ils concentrent l'ensemble des difficultés propres aux modèles très détaillés. Les ABS peuvent être considérés comme des modèles de simulation composés et pluri-formalisés ([DRQ04], [TDZ08]). V. Grimm et al. reconnaissent

1. Le mot « base » ici à le même sens que dans « base de données ».

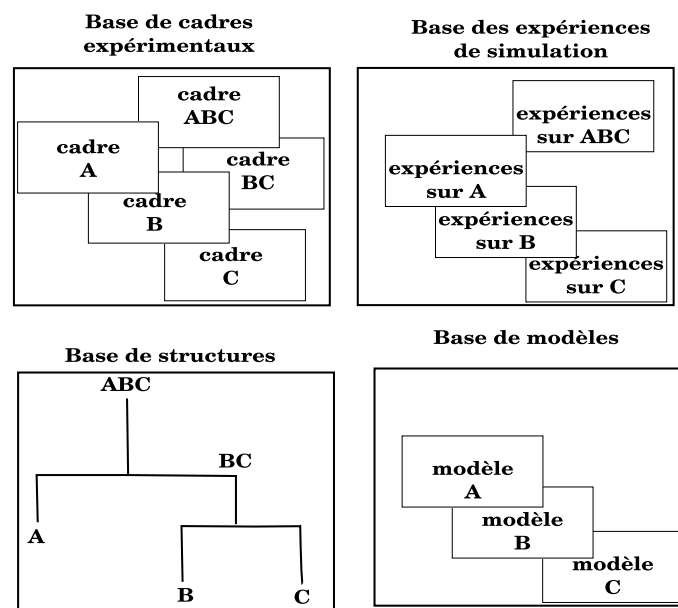


Figure 2 – Les quatre bases dans notre proposition (inspirées de [ZKP00] et [Aum07]). Ce découpage conserve une vision de composition hiérarchique des modèles et des cadre expérimentaux.

eux aussi implicitement que les modèles de type ABS sont des compositions de modèles élémentaires puisqu'ils définissent un élément « *sub-models* » qui décrit l'ensemble des composants de l'ABS dans le protocole ODD [GUD⁺10].

La méthode de communication et de spécification que nous proposons comprend cinq parties qui doivent être considérées dans l'ordre. La spécification complète d'un modèle composé se fait par réitération des cinq parties suivantes, avec possibilité d'omettre une partie si elle n'est pas nécessaire au niveau focal considéré :

- 1) Communication des objectifs de modélisation et du modèle.
- 2) Spécification formelle du cadre expérimental du modèle.
- 3) Spécification formelle du modèle.
- 4) Spécification formelle du simulateur et du coordinateur associé.
- 5) Communication des expériences de simulation.

Nous détaillons ces cinq parties dans ce qui suit.

4.2. Communication des objectifs de modélisation et du modèle

Cette partie est textuelle. Elle permet à toute personne qui veut utiliser un modèle de comprendre pour quels objectifs le modèle a été construit, quel est son contexte d'utilisation et quelles sont les hypothèses sur le système source qui ont guider les choix de conception. Nous utilisons le terme modèle dans ce qui suit pour désigner aussi bien un modèle composé qu'un modèle élémentaire. La fermeture sous couplage qu'exige notre méthode nous garantit qu'elle s'applique aux deux types de modèle.

4.2.1. Description générale des objectifs et du modèle

Pour structurer la description des objectifs, nous considérons tout d'abord la première partie (« *Overview* ») du protocole ODD [GUD⁺10]. Nous reprenons l'élément « *Purpose* » qui résume comment le modèle peut répondre à la question de recherche posée. L'élément « *Entities, state variables and scales* » devient « *Entities and scales* ». Il s'agit de décrire les entités du systèmes, leur résolution spa-

tiale et temporelle (à quelles échelles de temps et d'espace les entités sont-elles considérées). Cette partie doit également définir les niveaux focaux (hiérarchiques) du modèle en précisant à quel(s) niveau(x) les objectifs de modélisation (la question posée) seront étudiés. L'élément « *Process overview and scheduling* », est également modifié pour devenir « *Process overview* ». Il ne s'agit pas de décrire les algorithmes d'ordonnancement, ceux-ci devront apparaître dans la partie « spécification du simulateur » de notre méthode. Par contre il est toujours question de décrire les processus qui affectent les entités et de préciser si le modèle est en temps continu, discret ou hybride.

4.2.2. *Communication des hypothèses de construction du modèle*

Nous considérons ici la partie « *Design concepts* » de la dernière version du protocole ODD [GUD⁺10]. Dans cette partie, il y a originalement onze éléments à renseigner, nous considérons les dix premiers, à savoir « *basic principles, emergence, adaptation, objectives, learning, prediction, sensing, interaction, stochasticity and collectives* ». L'élément « *observation* » est inclut dans la section suivante. Comme il est dit dans le papier présentant ODD, l'ensemble des dix éléments ne doit pas être obligatoirement utilisé, seuls les éléments pertinents pour le modèle doivent apparaître.

4.2.3. *Communication des données en entrée et sortie du modèle*

Dans cette partie doit figurer la description des trajectoires d'états en sortie du modèle, c'est-à-dire calculées soient directement par le modèle, soit par des statistiques à partir des données simulées et utilisées pour évaluer le modèle au regard des objectifs, de la question posée. Ces variables d'états d'intérêt doivent être décrites par ce qu'elle représente dans le système source, leur nature discrète ou continue, leur unité, leur granularité et leur étendue. Les données (ou trajectoires d'état en entrée) doivent être décrite de la même manière que les trajectoires d'état en sortie.

4.3. *Spécification formelle du cadre expérimental du modèle*

Nous discutons dans cette partie, la manière de relier les objectifs de modélisation au cadre expérimental. Les objectifs de modélisation définissent les variables d'états d'intérêt. A partir de ces données, il est possible de définir les « *observables* » du modèle, c'est-à-dire une partie des données effectivement calculées et que l'on décide de stocker (ou de visualiser). Ces observables sont ensuite reliées aux variables d'état du modèle qui sont à leur tour reliées entre elles par la structure du modèle. Le cadre expérimental doit également spécifier le type des trajectoires d'états en entrée du modèle. Ces différentes trajectoires spécifient les différents contextes dans lequel le modèle peut être utilisé. Tous les cadres doivent être spécifiés formellement de la même manière. Pour cela, nous proposons l'utilisation du formalisme développé par Traoré et Muzzy [TM06], nous ne le reprenons pas ici et renvoyons le lecteur vers l'article original. L'important ici est de savoir que ce formalisme permet une spécification non ambiguë des cadres expérimentaux aux différents niveaux focaux.

4.4. *Spécification formelle du modèle*

Nous distinguons ici les modèles composés des modèles élémentaires. L'une ou l'autre des parties doit être utilisées en fonction de l'itération dans laquelle se situe la spécification.

4.4.1. *Spécification des modèles élémentaires*

Un modèle élémentaire est mono-formalisé. Il convient donc de décrire le modèle dans le formalisme approprié (événements discrets, équation différentielle, automates, réseaux de Petri, logique propositionnelle, etc.). Si le modèle ne peut être décrit formellement, il conviendra de donner les algorithmes de résolution. Les données en entrée et sortie du modèle doivent être identifiées, ainsi que les observables. Nous préconisons ici le formalisme DEVS (Discrete Event Systems specification) [ZKP00].

4.4.2. Spécification des modèles composés ou couplés

La composition des modèles couplés doit être décrite formellement en utilisant des structures qui décrivent le graphe de modèles couplés et les interfaces en entrée et sortie des modèles. Nous préconisons ici encore le formalisme DEVS pour les modèles couplés. Il est en effet nécessaire que le couplage vérifie la propriété de fermeture dans notre méthode, c'est le cas de DEVS.

4.5. Spécification formelle du simulateur et du coordinateur associé

Comme pour les modèles, les simulateurs sont divisés en deux types. Un type pour la simulation des modèles élémentaires et l'autre pour les modèles couplés, appelés alors coordinateurs. La figure 3 illustre la correspondance entre les hiérarchies de modèles et de simulateurs et coordinateurs. Les simulateurs échangent des événements via les coordinateurs. Le coordinateur racine joue le rôle de chef d'orchestre en envoyant les ordres d'exécution aux simulateurs, et donc en assurant la synchronisation.

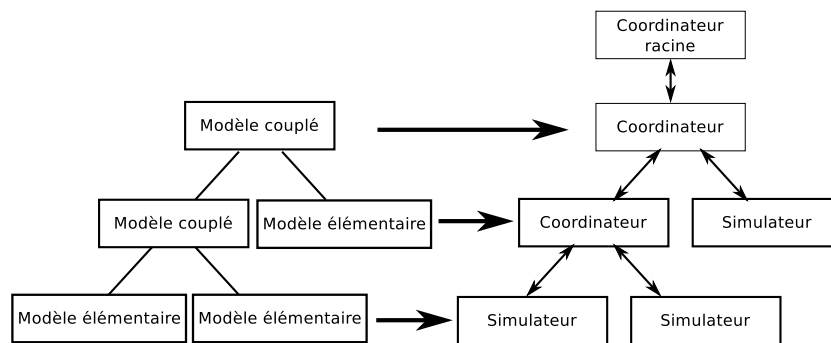


Figure 3 – Correspondance entre la hiérarchie de modèles élémentaires et de modèles couplés avec respectivement la hiérarchie des simulateurs et des coordinateurs associés (repris de Zeigler [ZKP00]). Les doubles flèches figurent les échanges d'événements qui existent entre les simulateurs via les coordinateurs.

De la même façon que pour la spécification des modèles, la spécification des simulateurs ou des coordinateurs doit être effectuée en fonction de l'itération dans laquelle se situe la spécification.

4.5.1. *Spécification des simulateurs de modèles élémentaires*

Pour spécifier les simulateurs de modèles élémentaires, la TMS introduit la notion de simulateur abstrait, qui décrit sous forme algorithmique un comportement dynamique. Par exemple les algorithmes d'intégration numérique des équations différentielles sont considérés comme des simulateurs abstraits dans le cadre de la TMS. Leur communication peut se faire sans ambiguïté. Les simulateurs abstraits définissent une ou plusieurs fonctions dont la signature est unique (le type et le nombre d'argument est unique). Les signatures des fonctions jouent le rôle d'interfaces avec l'extérieur du modèle élémentaire. L'ordre d'appel de ces fonctions est également unique et toujours le même (la suite de calculs effectués par un intégrateur d'équations différentielles ne peut pas être implémentée dans un ordre aléatoire et toujours différent).

4.5.2. *Spécification des simulateurs de modèles couplés*

Les informations échangées par les modèles sont gérées par un coordinateur qui correspond à la spécification opérationnelle du modèle couplé. Le coordinateur assure le routage des informations entre modèles de même niveau et entre modèles de niveau hiérarchiques $N + 1$ et $N - 1$. Au niveau le plus élevé, on trouve le coordinateur racine qui effectue l'appel des fonctions de chaque modèles composants dans un ordre déterminé. Le choix du composant imminent et le résultat d'un calcul fait au niveau des modèles élémentaires. Ainsi, l'avancement du temps est explicite au niveau modèle, le simulateur correspond à l'implémentation des choix fait par le modélisateur et n'impose pas de choix d'ordonnancement (voir [ZKP00] pour les détails). À chaque formalisme correspond un simulateur abstrait, le couplage des simulateurs abstraits réalise un simulateur multi-formalisme.

Nous n'entrons pas ici dans les détails de cette mécanique. Nous proposons de spécifier pour chaque modèle composant, un simulateur abstrait. Ainsi, les architectures et algorithmes des coordinateurs étant

connus, la description de l'ordonnancement du temps dans le modèle global devient transparente et non ambiguë. Il est défini au niveau des coordinateurs, indépendamment des modèles.

Cette spécification issue de la TMS permet de limiter les erreurs de correspondance entre modèle et simulateur. Le livre de Zeigler et al. [ZKP00] décrit notamment comment simuler de nombreux formalismes en implémentant des simulateurs abstraits en DEVS. Ainsi la preuve que le simulateur est homomorphe avec les modèles qu'il implémente peut être trouvée dans la littérature. Pour résumer, si le simulateur est spécifié et cette spécification référencée, alors il est possible de s'y référer pour toute implémentation future, ce qui allège la spécification complète des modèles.

4.6. Communication des expériences de simulation

Une expérience met en relation les bases de cadres expérimentaux, de structure et de modèle. Elle est stockée dans la base expérience de simulation de la figure 2. Une expérience utilise une structure, donne des valeurs aux paramètres et aux états initiaux des modèles en respectant les cadres expérimentaux associés et simule le modèle composé pour une durée définie (donnée a priori ou calculée par le modèle). Un plan d'expérience correspond bien à la génération d'expériences successives selon une politique donnée [STCR04]. Nous pensons tout comme Grimm et al. [GUD⁺10] que le plan d'expériences doit être dissocié de la présentation du modèle et faire l'objet d'une partie indépendante de la description du modèle.

Une fois que les quatre bases présentées figure 2 sont spécifiées, il est possible selon nous de discuter de façon beaucoup plus claire et crédible des résultats observés lors des plans d'expériences qui vont associer des structures, des modèles et des cadres expérimentaux. La façon de décrire ces plans d'expériences est libre néanmoins, elle doit être complète (description des plages de valeurs prises par les paramètres, graine du générateur aléatoire, valeurs initiales des variables d'états, références à la base de cadres expérimentaux et à la base de structures).

La figure 4 fait le lien entre les quatre bases de la figure 2 et les cinq points de la méthode que nous proposons. Une fois les différents points publiés pour un niveau focal donné, il est possible de se référer aux différentes bases pour obtenir les spécifications formelles et opérationnelles des modèles. La composition de modèles nécessite de repasser par l'ensemble des cinq points uniquement si l'on ajoute de nouveaux modèles élémentaires au nouveau niveau focal.

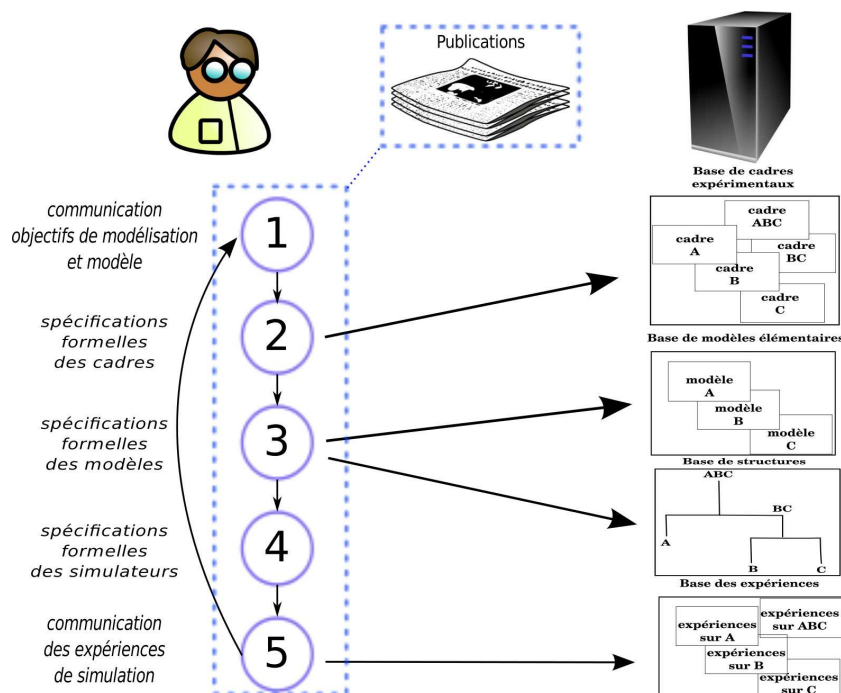


Figure 4 – Correspondance entre l'itération formée par les cinq points de la méthodologie et les quatre bases de notre proposition. Seuls les points 2, 3 et 5 donnent lieu à un enregistrement dans une base qui permet la réutilisation concrète des spécifications. L'ensemble des points doit faire l'objet de publication avant que les bases soient renseignées.

5. Discussion

Ce qui a motivé ce travail est un essai de rapprochement de deux communautés qui s'ignorent presque totalement, nous les avons appelées communauté ABS et TMS dans ce papier. Une des grandes différences entre ces communautés concerne les objectifs de modélisation. Il n'y a bien sûr pas de limite de séparation nette entre les deux communautés, néanmoins nous pouvons dire que la communauté ABS va davantage être orientée vers la caractérisation, la description et l'explication des processus qu'elle modélise de manière à augmenter la connaissance sur ces processus, à faire des prédictions, etc., alors que la communauté TMS va plutôt essayer de reproduire le comportement des systèmes qu'elle modélise à des fins d'optimisation et de conception. Ces deux usages de la modélisation peuvent être basés sur les mêmes outils de conception et de communication des modèles de simulation. C'est l'idée que nous défendons ici. De plus, ces deux usages se confondent lorsqu'il s'agit d'optimiser des décisions de gestion de systèmes socio-environnementaux par exemple, ce qui est une demande forte du côté simulation actuellement. Nous pensons que la méthode proposée ici permet d'intégrer ces deux usages de façon rigoureuse, dans une même spécification.

Notre proposition considère donc deux niveaux dans la communication des modèles. Le premier est le positionnement du modèle dans son contexte d'utilisation (question de recherche, d'ingénierie) en explicitant les concepts, les hypothèses méthodologiques qui justifient la conception du modèle. Le deuxième niveau concerne la spécification formelle des modèles. Ces deux niveaux sont naturellement interdépendants. La communauté d'utilisateurs de modèles ABS nous semble avoir répondu au premier niveau. Tandis que la communauté en TMS nous semble bien répondre au second.

Nous ne proposons pas une méthode pour tester la validité du modèle au regard de la question de recherche posée, mais bien une méthode de communication, de spécification des modèles de simulation à des fins de vérification. Nous n'avons volontairement pas présenté les aspects formels de notre méthode. Nous pensons qu'une description plus textuelle permet d'ouvrir notre réflexion à l'ensemble des utilisateurs de

la modélisation et de la simulation qui ne sont pas forcément des spécialistes de sa théorisation. Néanmoins, l'ensemble des aspects formels qui soutiennent notre méthode sont accessibles dans la littérature citée. Bien sûr, le travail de formalisation complet de notre méthode, présenté dans un seul document, devra être fait pour qu'elle puisse être d'abord testée, puis utilisée.

La relation entre le cadre expérimental et la question posée au modèle (les objectifs) est au cœur de notre méthode. C'est ce lien rendu explicite qui permet la communication rigoureuse du contexte d'utilisation du modèle et qui permet, avec la spécification du modèle lui-même, sa vérification et son utilisation par des tiers. Nous associons donc une base de cadre expérimentaux à la base de structures et à la base de modèles. Notre méthode se base sur une composition hiérarchique de cadre expérimentaux, comme proposé par G. Aumann [Aum07]. Un travail formel reste à faire à ce niveau car Aumann ne propose pas cette composition avec son formalisme. Néanmoins, la solution proposée par Traore et Muzy [TM06] peut servir de base à une telle formalisation. Dans ce contexte, nous pensons que notre méthode ouvre également des perspectives intéressantes sur l'automatisation de la construction de modèles couplés et la vérification de la validité du couplage à des niveaux sémantiques élevés. Des bibliothèques de cadres expérimentaux existent déjà et sont mises à disposition par des outils spécifiques [QDR09], ou même sur internet [CW09]. Pour l'instant, les implémentations dépendent des outils et il n'y a pas de moyen standard de les communiquer malgré les bases formelles qui les sous-tendent.

Dans notre proposition, la spécification du modèle global consiste en une structure choisie dans la base des structures possibles (telles que présentée figure 2) associée à la spécification des modèles composants. Cette solution nous permet de structurer notre modèle par niveaux hiérarchiques sur lesquels nous pouvons nous poser des questions spécifiques à travers des cadres expérimentaux qui sont définis pour chaque niveau. C'est là un apport important de la spécification des modèles complexes puisqu'en précisant l'argumentation sur un niveau hiérarchique, elle augmente la confiance que l'on a à propos des réponses formulées à ce niveau.

Cet article présente rapidement la notion de simulateur abstrait. Ces derniers sont écrits sous forme de pseudo-code. Le pseudo-code n'est pas exempt d'ambiguïté. Prenons un exemple très simple, l'écriture d'une boucle, nous pouvons proposer deux pseudo-codes qui produisent le même résultats :

```
pour tout élément i de la liste
  afficher i
fin pour

pour i = 1 jusqu'à la longueur de la liste
  afficher le ième élément de la liste
fin pour
```

Le deuxième exemple peut être considéré comme plus précis que le premier, car il laisse moins de liberté quant au choix d'implémentation pour l'algorithme de parcours de la liste. Il est possible de laisser plus ou moins de liberté au programmeur. Néanmoins dans les deux cas, les transitions d'états au niveau modèle sont homomorphiques avec celles du niveau simulateur abstrait, c'est cet homomorphisme qu'il faut prouver.

La dernière étape non discutée dans notre proposition est l'implémentation des simulateurs abstraits dans un langage particulier. Les techniques de vérifications qui s'appliquent ici sont celles du génie logiciel. L'ultime transformation du modèle en code exécutable est la phase de compilation ou d'interprétation du code. Cette phase est considérée comme non ambiguë car très standardisée (normes ISO), même si des choix d'optimisation, sur la représentation des réels par exemple, peuvent rester obscures et faire varier les résultats de calculs sur les flottants.

6. Conclusion

Notre proposition méthodologique pour la spécification de modèles de simulation de systèmes complexes repose principalement sur les travaux de B. Zeigler [ZKP00], G. Aumann [Aum07], Traore et Muzy

[TM06] et V. Grimm et al. [GUD⁺10]. Notre méthode peut être considérée comme un essai de synthèse et d'intégration de ces travaux. Il est encore nécessaire de l'affiner et de la tester. Nous avons commencé à implémenter cette méthode dans l'environnement de modélisation et simulation « *Virtual Laboratory Environment* » (VLE) [QDR09]. Toutefois, sa complexité limite encore son usage et un effort de clarification doit être entrepris.

La définition d'une méthode rigoureuse et partagée de construction de modèles pour l'étude des systèmes complexes est en construction depuis au moins une quarantaine d'année. Le livre de G. Braziller « *Hierarchy Theory, The Challenge of Complex Systems* » de 1973 est un bon exemple des débats qui animaient déjà les scientifiques de l'époque à ce sujet [Bra73]. La première phrase de la préface du livre de B. Braziller est la suivante :

« *We are well into a decade in which man can no longer evade the responsibility for his own survival. [...] It is a question of survival [...] in a society dominated by increasingly complex local constraints, but lacking stability and rational control.* ».

Trente huit ans plus tard, l'urgence nous impose qu'un protocole de spécification des modèles complexes soit finalisé. Ce protocole reste à écrire dans les détails. Nous espérons que cet article est un pas de plus dans la bonne direction.

Références

- [Aum07] G.A. Aumann. A methodology for developing simulation models of complex systems. *Ecological Modelling*, 202 :385–396, April 2007.
- [BCK09] J. H. Byun, C. B. Choi, and T. G. Kim. Verification of the devs model implementation using aspect embedded devs. In *Proceedings of the 2009 Spring Simulation Multiconference*, San Diego, USA, 2009.

- [Bra73] G. Braziller. *Hierarchy Theory, The Challenge of Complex Systems*. Howard H. Pattee, 1973.
- [CW09] Rachid Chreyh and Gabriel Wainer. Cd++ repository : an internet based searchable database of devs models and their experimental frames. In *Proceedings of the 2009 Spring Simulation Multiconference, SpringSim '09*, pages 159 :1–159 :8, San Diego, CA, USA, 2009. Society for Computer Simulation International.
- [DRQ04] R. Duboz, E. Ramat, and G. Quesnel. Systèmes multi-agents et théorie de la modélisation et de la simulation : une analogie opérationnelle. In *Actes des douzièmes Journées Francophones sur les Systèmes Multi-Agents (JF-SMA) - Systèmes multi-agents défis scientifiques et nouveaux usages*. Olivier Boissier et Zahia Guessoum eds., 2004.
- [Fei54] J.K. Feibleman. Theory of integrative levels. *Br. J. Philos. Sci.*, 5(17) :59–66, 1954.
- [GUD⁺10] V. Grimm, Berger U, DeAngelis DL, Polhill G, Giske J, and Railsback SF. The odd protocol : a review and first update. *Ecological Modelling*, 221 :2760–2768, 2010.
- [Hah08] C. Hahn. A domain specific modeling language for multiagent systems. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume I. AAMAS*, 2008.
- [Kli85] G.J. Klir. *Architecture of Systems Complexity*. Saunders, New York, 1985.
- [LW05] Y. Labiche and G. Wainer. Towards the verification and validation of devs models. In *in Proceedings of 1st Open International Conference on Modeling and Simulation*, pages 295–305, 2005.
- [PTW07] L. Padgham, J. Thangarajah, and M. Winikoff. Auml protocols and code generation in the prometheus design tool.

- In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems (AAMAS)*, New York, USA, 2007.
- [QDR09] Gauthier Quesnel, Raphaël Duboz, and Éric Ramat. The Virtual Laboratory Environment – An operational framework for multi-modelling, simulation and analysis of complex dynamical systems. *Simulation Modelling Practice and Theory*, 17 :641–653, April 2009.
- [SDS10] A. Sturm, D. Dori, and O. Shehory. An object-process-based modeling language for multiagent systems. *Systems, Man, and Cybernetics, Part C : Applications and Reviews*, 40(2) :227–241, March 2010.
- [STCR04] Andrea Saltelli, Stefano Tarantola, Francesca Campolongo, and Marco Ratto. *Sensitivity Analysis in Practice : A Guide to Assessing Scientific Models*. Halsted Press, New York, NY, USA, 2004.
- [TDZ08] J.P. Treuil, A. Drogoul, and J.D. Zucker. *Modélisation et simulation à base d’agents. Exemples commentés, outils informatiques et questions théoriques*. DUNOD, 2008.
- [TM06] Mamadou K. Traore and Alexandre Muzy. Capturing the dual relationship between simulation models and their context. *Simulation Modelling Practice and Theory*, 14 :126–142, 2006.
- [Var06] F. Varenne. *Les notions de métaphore et d’analogie dans les épistémologies des modèles et des simulations*. Petra, 2006.
- [VUF⁺06] Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, Goss-Custard J, Grand T, Heinz S, Huse G, Huth A, Jepsen JU, Jørgensen C, Mooij WM, Müller B, Pe’er G, Piou C, Railsback SF, Robbins AM, Robbins MM, Rossmanith E, Rüger N, Strand E, Souissi S, Stillman RA, Vabø R, Visser U, and DeAngelis DL. A standard protocol for

- describing individual-based and agent-based models. *Ecological Modelling*, 198 :115–126, 2006.
- [Wai10] G. Wainer. *Discrete-event Modeling and Simulation : Theory and Applications*. CRS Press Inc, 2010.
- [Wil99] U. Wilensky. Netlogo. center for connected learning and computer-based modeling, northwestern university. evans-ton, il., 1999.
- [Zei76] B. P. Zeigler. *Theory Of Modeling and Simulation*. Wiley Interscience, 1976.
- [ZKP00] B. P. Zeigler, D. Kim, and H. Praehofer. *Theory of modeling and simulation : Integrating Discrete Event and Continuous Complex Dynamic Systems*. Academic Press, 2000.

Annexe H

Résultats de simulation

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	test ^a
f1	1	178070	178070	70.3341	2.667e-16	***
f2	2	51067	25534	10.0852	4.795e-05	***
f3	1	0	0	4.937e-06	0.9982	
f4	1	492	492	0.1942	0.6596	
Residuals	714	1807685	2532			

TABLE H.1: Analyse de la variance du critère c_1 en fonction des critères f_1 à f_4 .

a. Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Estimate	Std.Error	t value	Pr(> t)	Test ^a
(Intercept)	100.643056	4.593267	21.911	< 2e-16	***
f1 0.045	31.452778	3.750387	8.387	2.67e-16	***
f2 4	2.187500	4.593267	0.476	0.634	
f2 5	18.858333	4.593267	4.106	4.50e-05	***
f3 0.4	-0.008333	3.750387	-0.002	0.998	
f4 0	1.652778	3.750387	0.441	0.660	

TABLE H.2: Coefficients modèle linéaire $c_1 \sim f_1 + f_2 + f_3 + f_4$

a. Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Estimate	Std.Error	t value	Pr(> t)	Test ^a
(Intercept)	60.544	1.749	34.625	<2e-16	***
f1 0.045	8.219	1.428	5.757	1.27e-08	***
f2 4	-14.576	1.749	-8.336	3.94e-16	***
f2 5	-19.098	1.749	-10.922	<2e-16	***
f3 0.4	-14.334	1.428	-10.040	<2e-16	***
f4 1	-1.055	1.428	-0.739	0.46	

TABLE H.3: Coefficients modèle linéaire $c_3 \sim f_1 + f_2 + f_3 + f_4$

a. Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Estimate	Std.Error	t value	Pr(> t)	Test ^a
(Intercept)	51.9028	2.4996	20.764	< 2e-16	***
f1 0.045	19.0164	2.0409	9.318	< 2e-16	***
f2 4	-0.4402	2.4996	-0.176	0.86026	
f2 5	6.9792	2.4996	2.792	0.00538	**
f3 0.4	-0.1647	2.0409	-0.081	0.93570	
f4 1	0.5764	2.0409	0.282	0.77771	

TABLE H.4: Coefficients modèle linéaire $c4 \sim f1 + f2 + f3 + f4$

^a. Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Annexe I

Exemples d'utilisation du modèle de système épidémiologique

I.1 Simulation de la propagation de la grippe aviaire hautement pathogène dans le réseau de pro- duction du Nord Vietnam

Simulation of avian influenza propagation in the poultry production network of the North Vietnam

Raphaël Duboz, Bruno Bonté, Stéphanie Desvaux, Vladimir Grosbois

Contact: raphael.duboz@cirad.fr



UR AGIR
Animal et Gestion Intégrée des Risques
TA C 22E Campus International de Balaruc
34398 Montpellier cedex 5
France

Highly Pathogenic Avian Influenza (HPAI), subtype H5N1, is now endemic in the Red River Delta, North Vietnam. Surveillance and control of the virus circulation in poultry remain a challenge. One shared hypothesis is that poultry trading network drives the spread of the virus at the scale of the Delta.

Interviews were conducted in a sub-area of the Red River Delta, Vietnam, in 2008 in order to perform a Social Network Analysis (SNA). A representative sample of semi-commercial farms and backyards poultry producers was interviewed in 6 communes (Figure 1) and their trading connections with poultry markets and traders were recorded. The majority of the exchanges are performed by traders. As illustrated by Figure 2, the poultry trading network is highly unconnected if we don't consider the traders' activity. Unfortunately, traders identified by this first survey were very difficult to trace. Therefore, another survey was performed on a representative sample of the population of traders and provided informations at the commune level. We know the communes in which a trader sells and/or buys to markets, backyards or farms. Nevertheless, we can not identify those markets, backyards or farms.

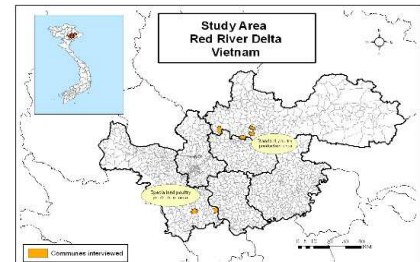


Figure 1: Localization of interviewed communes (orange areas). The two types of poultry production were selected (standard and specialized) to be representative of the Red River Delta production basin.

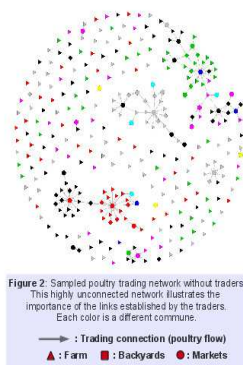


Figure 2: Sampled poultry trading network without traders. This highly unconnected network illustrates the importance of the links established by the traders. Each color is a different commune.

CONTEXT
How essential is it to capture trader's activities at the production unit level?

MODELING QUESTION
What is the impact of the infra-communal network structure on the disease spread at the global level?

We build a simulation model of the avian influenza propagation in a representative poultry production network.

We generated networks representing traders activities at the intra-communal level based on the traders interviews.

We perform an experimental design to test the influence of those networks on the final size of the epidemics.

THE MODEL

We model the spread of a disease in a graph. Nodes are either farms, backyards or markets. Links are oriented and represent poultry flows between nodes. The modeled network is a graph of timed finite state automata.

Nodes: Each node N is specified with the following set of state:

$$N = \{S, E, I, R\}$$

with the following set of transitions:

$$T = \{S \rightarrow E, E \rightarrow I, I \rightarrow R\}$$

and the following associated life spans:

$S = \text{infinity}$; $E = \text{latent period}$; $I = \text{infectious period}$; $R = \text{infinity}$

Poultry flow dynamic: Every time step, each node sends an event to the nodes it is connected with, with a probability P . This probability is defined as the inverse of the frequency of sells between two connected nodes.

Infection: The transition between S and E occurs when an event comes from one infected connected node.

EXPERIMENTAL DESIGN

We test the influence of 3 factors (Table 1)

- sells frequency
- network density (traders connectivity)
- infra-communal structure

Time step is one day and simulated period is 1 month. We perform 15 replicas for each factor combination.

Table 1: List and values of the factors tested in the experimental design. Due to the stochastic characteristic of the model, we computed 15 simulations for each factors combination. 2250 simulations were computed. Tested values were obtained by expert interview.

Factors	Modalities	N
Frequency of sells to one linked node	{6, 8, 10, 12, 14} (number of sells per year)	5
Trader-connectivity	- Low connectivity (20 to 30 farms and 3 to 5 markets) - High connectivity (30 to 40 farms and 6 to 10 markets)	2
Randomly generated graph	Randomly generated according to 15 traders' survey information on trading between communes (Figure 4)	15

SIMULATION RESULTS

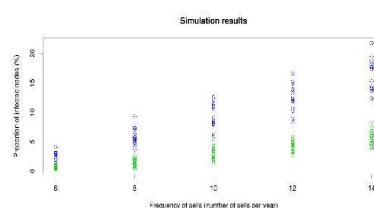


Figure 3: negative binomial general linear model fitted on simulation results.

For each factors' combination the chart shows the proportion of nodes which have been infected during the simulation (y axis). Color stands for the type of connectivity (blue-high connectivity, green-low connectivity). The x axis stands for the frequency of sells factor value. The modalities of the 'randomly generated graph' factor are not identified.

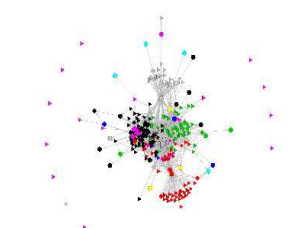


Figure 4: Example of a generated poultry trading network with traders. This network has been randomly generated according to knowledge on inter-communal trade (same legend as Figure 2).

ANALYSIS

Table 2 provides the main results of an analysis of variance. The objective is to quantify the impact of the three factors modalities on the final size of the epidemic. As shown by table 2, the trader connectivity density and frequency of sells are highly significant regarding the final size of the simulated epidemic. The interaction between these two factors is less important, even if it is significant. The different graph topologies generated are not significant.

Table 2: Main result of the analysis of variance

Factor	Mean Square Sum	F Value
Frequency of sells	130.4	68.75***
Trader connectivity	402.9	212.4***
graph	2.4	1.26
Trader connectivity and frequency of sells interaction	5.8	3.06*

CONCLUSION

This preliminary study indicates that the identification of the traders clients and suppliers at the individual level is not necessary to accurately reproduce an epidemic in the studied network. The knowledge of the type and the number of clients and suppliers in each commune, as well as the frequencies of interactions is sufficient.



References:
Munoz-Gonzalez and Michon J. J. 2009. Epidemic threshold and network structure: The interplay of probability of transmission and of persistence in small size directed networks. *Ecological Complexity*, 5(1): 1-8, 2009

GRIPAVI PROJECT
Funded by French Ministry of Foreign and European Affairs



Model design and execution

http://Me.toulouse.inra.fr/

I.2 Simulation de méthodes de captures-recaptures comme un nouvel outil d'évaluation de la sur- veillance en santé animale

Simulation of capture-recapture methods as a new tool to assess animal disease surveillance :

Application to FMD outbreaks reports in Cambodia

T. Vergne^{1&2}, B. Dufour², B. Bonté¹, S. San³, T. Sothyra⁴, A. Boucho⁵, F. Roger¹, V. Grobois¹

Background

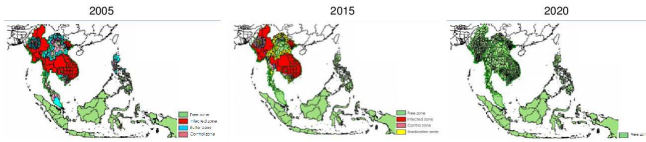


Figure 1: SEAFMD objective is eradication of FMD by 2020 with a zoning approach

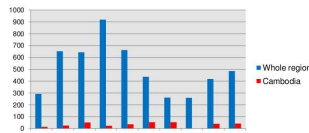


Table 1: Number of outbreaks declared to the SEAFMD campaign (2000-2009)

Are CR methods relevant to assess the national reporting systems ?

Capture-recapture methodology in a few words

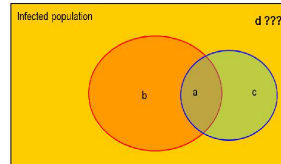


Table 2: Representations of a two sources capture-recapture analysis

	Source 1		
	yes	no	
Source 2	yes	a	c
	no	b	d ?
		N ₁	N

Source independence

$$P(1 \& 2) = P(1) * P(2)$$

$$a/N = (a+b)/N * (a+c)/N$$

Estimated infected population size

$$N_{LP} = (a+b)*(a+c)/a \text{ (Lincoln-Petersen estimator)}$$

Estimated sensitivity of source 1

$$Sb1 = (a+b)/N_{LP}$$

Underlying assumptions

Population is **closed** (verified)

Cases detected are **real cases** (verified)

Sources are **locally independent** (objective of the 2nd source design)

There is **no heterogeneity** of capture among individuals (to be discussed)

Field study definition

Case definition

Village with clinical cases

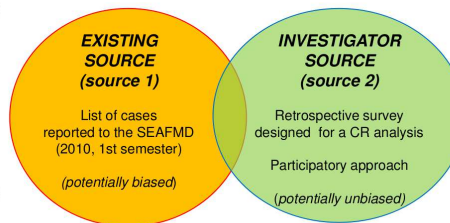
Place

Pursat and Kampong Speu

Time period

1st semester of 2010

Material



Investigator source Sampling strategy ?

Sample size

Random approach
Snowball approach
(=after an initial random sampling phase, sample the units linked with the units sampled in the first phase and detected as infected)

Simulation as a tool to evaluate the best sampling strategy for the second source

How to create the more efficient second source to get a good estimation ?

Dynamic simulation of an epizootic of an infectious disease in a scale-free graph

We simulated a very simple epizootic of an infectious disease through a scale-free graph which nodes represent the **unit of interest** (villages here), and which links represent the **epidemiological connections** between the nodes (like traders, markets or just spatial proximity).

At the end of the epidemic, we have a final pattern on which we test the **different sampling strategies** for the investigator source. Each strategy will be compared to the other ones in term of bias and coefficient of variation (CV) of the calculated estimators.

Figure 2: Representation of the nodes at the end of the simulated epidemic

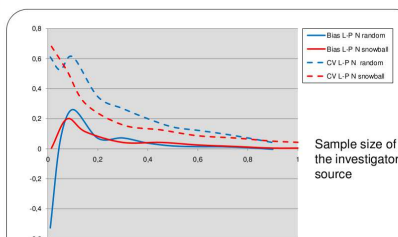
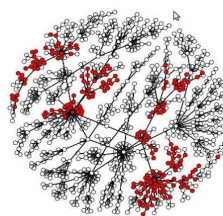


Figure 3: Influence of the sample size of the investigator source (source 2) on the L-P estimator of the number of infected nodes for a random and a snowball approach.

Prevalence = 0,2; Sensitivity of the existing source (source 1) = 0,2.

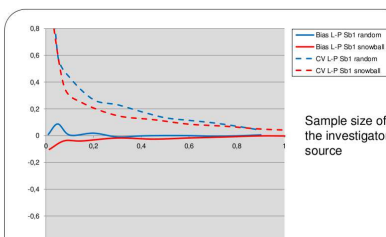


Figure 4: Influence of the sample size of the second source (size 2) on the L-P estimator of the sensitivity of the source 1 for a random and a snowball approach.

Prevalence = 0,2; Sensitivity of the existing source (source 1) = 0,2.

Discussion

The data presented here only deal with the **random approach** and the **snowball approach**.

	Random approach	Snowball approach
Accuracy	++	+
Precision	+	++
Cheapness	+	++
Best designed for...	-	Infected population size Sensitivity of the existing source

Table 3: Advantages and drawbacks of the random and the snowball approaches.

•The snowball sampling approach results in a higher precision at the expense of a lower accuracy.

•An important gain from the snowball approach is the **better ease of data collection and thus lower sampling costs**, which can be a great advantage in the context of developing countries like Cambodia.

•In our simulations, the cases detected by the first source were randomly attributed whereas for real epidemiological surveillance systems, **some nodes (big villages, commercial farms) are more easily detected than others**. Thus, the next step in our investigation will be to assess a sampling strategies for stratified populations.

➔ **Final objective: Define the best and more realistic strategy into the field in Cambodia to assess the provincial reporting system.**

¹ Centre de coopération internationale en recherche agronomique pour le développement (CIRAD), Montpellier, France

² EpiMAI, Ecole Nationale vétérinaire d'Alfort, Maisons-Alfort, France

³ National Veterinary Research Institute, Phnom Penh, Cambodia

⁴ Department of Animal Health and Production, Phnom Penh, Cambodia

⁵ OIE, World Organization of Animal Health, South East Asia sub-Regional Representation, Bangkok, Thailand

Annexe J

Utilisation d'un modèle de système épidémiologique couplé à un modèle de surveillance épidémiologique pour l'aide à la conception des systèmes de surveillance

*ANNEXE J. UTILISATION D'UN MODÈLE DE SYSTÈME
ÉPIDÉMIOLOGIQUE COUPLÉ À UN MODÈLE DE SURVEILLANCE
ÉPIDÉMIOLOGIQUE POUR L'AIDE À LA CONCEPTION DES SYSTÈMES
DE SURVEILLANCE*

A modular simulation tool to help designing epidemics surveillance: Work in Progress.

B. Bonté¹, R. Duboz¹

¹ CIRAD, AGIRs unit, Montpellier, France and AIT-CSIM laboratory, Bangkok, Thailand

*bruno.bonte@cirad.fr

Keywords: Modeling, Simulation, DEVS, Surveillance

Summary

There is nowadays a need of new surveillance systems, lighter, cheaper and quantitatively evaluated. Hence, new statistical tools are being developed and new surveillance systems are being designed. Then, raise the issues of evaluating these new surveillance systems and build a clear view of their implementation, shared by researchers, actors and decision makers. In order to address these issues, we are building a modular and multi-formalisms simulation tool that allows simulating both the disease spread, and its surveillance. This simulation tools is a framework that can be used in two ways: Firstly to visualize at the same time epidemic spread and surveillance on didactic interfaces; and secondly to perform intensive simulation plans in order to test surveillance protocols in different scenarios. This is a work in progress; the basic structure has been implemented using Discrete Event System specification and the Virtual Laboratory Environment software system. Future work will be to implement a case study and to introduce the disease control component, which is essential to surveillance evaluation.

Introduction

Designing new surveillance systems is a challenge. Sampling strategies, as well as data analysis must be optimized and evaluated but experiments on the field are very expensive and irreproducible. Furthermore, evaluation is an issue because we can't compare the result of surveillance with the unknown real epidemiological state. For these reasons we are building a simulation tool dedicated to surveillance systems designers. Contrarily to real experiments, computer simulations are cheap and reproducible. Moreover, we can know the real epidemiological state if we simulate it. Hence, this tool must allow simulating both, the **sampling performed by the surveillance system**, and the **epidemic spread**, in order to:

- **quantitatively evaluate** the adequacy between surveillance system estimation and real epidemic in **different scenarios**;
- optimize the **sampling process parameterization**;
- present the surveillance system sampling strategy in a didactic and/or contextualized way in order to discuss it with every actor of the surveillance.

Materials and methods

Our simulation tool simulates coupled models of disease spread and of surveillance monitoring. These models are specified using Parallel DEVS specification (1) and are implemented in the open source Virtual Laboratory Environment (VLE) software system (2).

The Parallel Discrete Event System Specification (DEVS) has two advantages. Firstly it allows multi-formalisms: we can specify dynamic systems models using different formalisms such as automaton or differential equations. Secondly, it is modular: A Parallel DEVS model (representing the global system) is composed of several Atomic DEVS models (representing sub-systems) coupled together. So we can take a sub-system and replace it by another one. Hence, we can simulate a model of an epidemics formalized as a differential equation system, coupled with a model of surveillance activity formalized as an automaton performing regular observations. Thus, we can keep the same model of epidemics but coupled with different models of surveillance in order to assess which surveillance gives the best results.

The VLE software system suite provides a set of tools to implement and simulate DEVS models. Beside other features, it provides an interface with R software system (3) that we use to explore and analyze models by running simulation plans (several simulations); and visualization plug-ins that we use to visualize the results of a single simulation, facilitating discussions.

Simulation tool presentation

Disease spread modular model: Literature in epidemiology proposes huge amount of epidemiological models (4). Obviously, using different models implies making different hypothesis and simplifications that will have an impact on simulation results. For that reason, we are building the surveillance package as a modular and scalable tool in order to both, be able to test different hypothesis and assess the impact of different simplifications.

However, we think that the disease spread models we consider must be spatial and dynamic to allow an accurate description of the surveillance system that will monitor it. Therefore, the disease spread model has a basic structure. It is built as a set of “places” linked together by an “infectious contact network”. The infectious contact network is represented by a directed graph either valued or

not if some infectious links are supposed to be stronger than others. The places are situated both in a 2D space and in the infectious contact network. The epidemic dynamics is then specified in the “places” models which are coupled together using the “infectious contact network” in order to constitute the disease spread model. Hence, we can test different disease spread models in the same network.

Surveillance modular model: The surveillance model is a set of surveillance components coupled with the disease spread model. We identified two different kinds of surveillance components. The first ones are event-based surveillance components that automatically receive information events any time the epidemiological state changes. And the second ones are pro-active surveillance components that must request for information events. A surveillance component is then specified by three features: an observation policy (what to observe and when?); a perception error (is the information damaged?); and an analyze feature (How informations are aggregated?).

Simulation Tool outputs

Visualization of a single experiment: In figures 1 and 2, we present two kinds of simulations outputs that can be produced by the simulation tool. In this toy simulation, a hypothetic disease spreads through the villages of the Phitsanuloc Province in North Thailand. Villages have an epidemiological state that can be susceptible, infectious or recovered. Then, an infectious contact network has been arbitrarily generated according to villages distances one to another (see figure 1). At the same time two surveillance components are monitoring the epidemics and estimating prevalence at a regular time step of 5 time units. The first one observes all villages but have a 0.25 probability of a false positive observation and the same probability of a false negative observation. The second one observes a random sample of 10% of all villages with a perfect observation (see figure 2). Hence, simulation outputs can be visualized both as dynamic maps, as shown figure 1, and as charts, as shown figure 2.

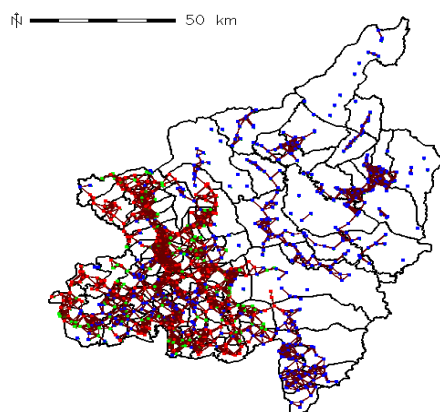


Figure 1: Epidemic spread output (snapshot during simulation). Visualization of disease spread in toy simulation where disease spreads in villages network. Villages (dots) colours indicates epidemiological state (Blue=Suceptible, Red= Infectious, Green=recovered)

Building and analyzing experimental plans: The main interest of our simulation tool is that it can be used to quantitatively evaluate sampling strategies. As we said, performing real experiments on the field is hardly possible. Using simulations allows us to repeat virtual experiments and hence to test many possible scenarios. Interface between VLE and R software gives us access to many statistical tools in order to explore parameters space and then to analyze simulation results. Experimental plans distribution feature of VLE allows us to perform intensive simulation plans on computer clusters.

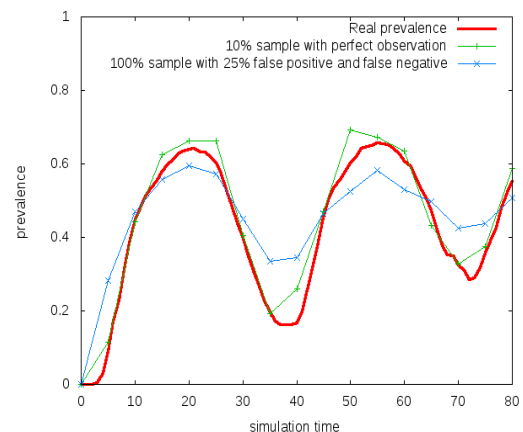


Figure 2: Surveillance monitoring visualization Visualization of prevalence monitored by different surveillance components during simulation of figure 1.

Future work

Implement a case study on HPAI in SEA: HPAI subtype H5N1, is nowadays endemic in several countries in eastern-south Asia. Surveillance of the virus circulation in poultry remains a challenge: Exporting countries must prove a minimal prevalence in their poultry production chain to be declared free of the virus. To do that, they must prove the quality of their surveillance system. At the same time, the surveillance system must be able to perform early detection and allow controlling disease spread. Classical surveillance systems are expansive and not efficient. Hence, there is a need of new surveillance systems, lighter, cheaper and quantitatively evaluated. In this case study we will evaluate if Capture-Recapture analysis (5,6,7) would have helped in the context of the 2004 epidemics of HPAI in northern Thailand (see map Figure 1.).

Add a model of disease control: A surveillance system is often associated to a disease control system. For two reasons we must consider the associated control system when evaluating a surveillance system. The first reason is that the surveillance system is often supposed to perform early detection in order to control the epidemics. Thus, surveillance evaluation must include an evaluation of the couple surveillance/control capacity to limit the spread of

the disease. The second reason is that an observation performed by a surveillance system will have an indirect feedback on the disease spread, through eventual control actions, thus the two systems “surveillance”/“disease spread” can't be considered separately but must be simulated as two coupled systems in interaction through a control system. We will now work on the introduction of a control model component in our simulation tool.

Conclusion

We built a general framework of a generic and modular simulation tool meant to help surveillance system design. Designers can use this tool to evaluate quantitatively new sampling strategies and to present and discuss these new strategies using didactic visualization features such as dynamic maps.

The tool is still at a prototype stage but the modeling formalism (DEVs), as well as the open source software used for its development make it modular and scalable. It means that any case study developed with it from now on will participate to improve it and complete it.

Acknowledgment

The authors want to thank Mathilde Paul, Timothée Vergne and Flavie Goutard who are working on the first case study, in the framework of the DGAI-funded FRIA-08-009 REVASIA project. They hence participated to the design of the simulation tool. Many thanks also to the VLE software development team who made this work possible and helped for implementation.

References

- (1) Zeigler et al., 2000. Academic Press.
- (2) Quesnel et al., 2009. Simulation Modelling Practice and Theory, **17**: 641–653.
- (3) R Development Core team, 2009. <http://www.r-project.org>.
- (4) M. J. Keeling and P. Rohani, 2008. Princeton University Press.
- (5) Hook and Regal, 1995. *Epidemiol Rev.* **17**: 243-64.
- (6) IWGDMF, 1995. *Am J Epidemiol.* **142**: 1047-58.
- (7) Vergne et al., 2010. *13th AITVM conference*: 31-33.